A globin domain in a neuronal transmembrane receptor of caenorhabditis elegans and ascaris suum : molecular modeling and functional properties

Tilleman, Lesley ; Germani, Francesca ; De Henau, Sasha ; Helbo, Signe ; Desmet, Filip ; Berghmans, Herald ; Van Doorslaer, Sabine ; Hoogewijs, David ; Schoofs, Liliane ; Braeckman, Bart P ; Moens, Luc ; Fago, Angela ; Dewilde, Sylvia

Abstract: We report the structural and biochemical characterization of GLB-33, a putative neuropeptide receptor that is exclusively expressed in the nervous system of the nematode Caenorhabditis elegans. This unique chimeric protein is composed of a 7-transmembrane domain (7TM), GLB-33 7TM, typical of a G-protein-coupled receptor, and of a globin domain (GD), GLB-33 GD. Comprehensive sequence similarity searches in the genome of the parasitic nematode, Ascaris suum, revealed a chimeric protein that is similar to a Phe-Met-Arg-Phe-amide neuropeptide receptor. The three-dimensional structures of the separate domains of both species and of the full-length proteins were modeled. The 7TM domains of both proteins appeared very similar, but the globin domain of the A. suum receptor surprisingly seemed to lack several helices, suggesting a novel truncated globin fold. The globin domain of C. elegans GLB-33, however, was very similar to a genuine “myoglobin”-type molecule. Spectroscopic analysis of the recombinant GLB-33 GD showed that the heme is pentacoordinate when ferrous and in the hydroxide-ligated formwhenferric, even at neutral pH. Flash-photolysis experiments showed overall fast biphasic CO rebinding kinetics. In its ferrous deoxy form, GLB-33GDis capable of reversibly binding O2 with a very high affinity and of reducing nitrite to nitric oxide faster than other globins. Collectively, these properties suggest that the globin domain of GLB-33 may serve as a highly sensitive oxygen sensor and/or as a nitrite reductase. Both properties are potentially able to modulate the neuropeptide sensitivity of the neuronal transmembrane receptor.

DOI: https://doi.org/10.1074/jbc.M114.576520

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 290, NO. 16, pp. 10336–10352, April 17, 2015

A Globin Domain in a Neuronal Transmembrane Receptor of *Caenorhabditis elegans* and *Ascaris suum*

*MOLECULAR MODELING AND FUNCTIONAL PROPERTIES**

Received for publication, April 25, 2014, and in revised form, February 7, 2015 Published, JBC Papers in Press, February 9, 2015, DOI 10.1074/jbc.M114.576520

Lesley Tilleman[‡], Francesca Germani[‡1], Sasha De Henau[§1], Signe Helbo[¶2], Filip Desmet[‖], Herald Berghmans[‡], Sabine Van Doorslaer[‖], David Hoogewijs[‡‡], Liliane Schoofs[§§3], Bart P. Braeckman[§], Luc Moens[‡], Angela Fago[¶2,4], and Sylvia Dewilde[‡4,5]**

From the Departments of [‡]Biomedical Sciences and [‖]Physics, University of Antwerp, 2610 Antwerp, Belgium, the [§]Department of Biology, Ghent University, 9000 Ghent, Belgium, the [¶]Department of Bioscience, Aarhus University, 8000 Aarhus, Denmark, the [**]Institute of Physiology and Zürich Center for Integrative Human Physiology, University of Zürich, 8006 Zürich, Switzerland, [‡‡]Institute of Physiology, University of Duisburg-Essen, D-45147 Essen, Germany, and the [§§]Functional Genomics and Proteomics Group, KU Leuven, 3000 Leuven, Belgium

Background: GLB-33 is a putative neuropeptide receptor in *C. elegans* with a globin domain.
Results: Recombinant globin domain displays a ferric hydroxide-ligated form. When reduced, it can bind CO or O_2 and reduce nitrite to NO.
Conclusion: The globin domain may serve as an oxygen sensor or nitrite reductase.
Significance: Oxygen-sensing mechanisms are relevant for neuropeptide receptor binding.

We report the structural and biochemical characterization of GLB-33, a putative neuropeptide receptor that is exclusively expressed in the nervous system of the nematode *Caenorhabditis elegans*. This unique chimeric protein is composed of a 7-transmembrane domain (7TM), GLB-33 7TM, typical of a G-protein-coupled receptor, and of a globin domain (GD), GLB-33 GD. Comprehensive sequence similarity searches in the genome of the parasitic nematode, *Ascaris suum,* revealed a chimeric protein that is similar to a Phe-Met-Arg-Phe-amide neuropeptide receptor. The three-dimensional structures of the separate domains of both species and of the full-length proteins were modeled. The 7TM domains of both proteins appeared very similar, but the globin domain of the *A. suum* receptor surprisingly seemed to lack several helices, suggesting a novel truncated globin fold. The globin domain of *C. elegans* GLB-33, however, was very similar to a genuine myoglobin-type molecule. Spectroscopic analysis of the recombinant GLB-33 GD showed that the heme is pentacoordinate when ferrous and in the hydroxide-ligated form when ferric, even at neutral pH. Flash-photolysis experiments showed overall fast biphasic CO rebinding kinetics. In its ferrous deoxy form, GLB-33 GD is capable of reversibly binding O_2 with a very high affinity and of reducing nitrite to nitric oxide faster than other globins. Collectively, these properties suggest that the globin domain of GLB-33 may serve as a highly sensitive oxygen sensor and/or as a nitrite reductase. Both properties are potentially able to modulate the neuropeptide sensitivity of the neuronal transmembrane receptor.

Nematodes often inhabit low oxygen environments where parameters such as food supply, temperature, and oxygen pressure vary. These fluctuations force the nematode to adapt its behavior to survive (1). Globins can play a role in this adaptation. The globin family is a heterogeneous group of proteins sharing the globin fold and burying a heme cofactor in a hydrophobic pocket, where diatomic gaseous ligands like O_2 can bind to the heme iron atom. *In silico* analysis of the genome of the free-living nematode *Caenorhabditis elegans* resulted in the identification of a family of 33 globin genes, labeled *glb-1* to *glb-33*, that are all transcribed. The encoded globins, GLB-1 to GLB-33, are localized in a specific subset of cells (2–5). Characterization of some of these globins indicated that they are involved in a wide variety of cellular processes. For instance, GLB-1 may function in regulating facilitated O_2 diffusion in a small subset of cells (6). GLB-6 serves a role as redox sensor responding to oxidative stress, either directly or indirectly by the change in O_2 levels (7). GLB-5 is an O_2 sensor that interacts with the NPR-1 receptor as part of a foraging strategy (8, 9). GLB-26 is a myristoylated redox protein, involved in the defecation cycle under stress conditions (6, 10). The role of all other globins in *C. elegans* remains unknown. One of the *C. elegans* globin genes, *glb-33*, is expressed in neuronal cells and has been predicted to code for a 542-amino acid-long chimeric protein, GLB-33.

The N-terminal domain of GLB-33 (residues 1–372) is composed of 7 hydrophobic α-helices and represents a 7-transmembrane domain (GLB-33 7TM[6]) typical of a G-protein-cou-

* This work was supported in part by Bijzonder Onderzoeksfonds TOP University of Antwerp 2006 (to S. D. and L. M.) and by Fund for Scientific Research Project G.0247.09 (to S. D., L. M., S. V. D., and B. B.).
[1] Ph.D. fellows of the Fund for Scientific Research.
[2] Supported by the Lundbeck Foundation and The Danish Council for Independent Research, Natural Sciences Grant 10-084565.
[3] Supported by KU Leuven and the Fund for Scientific Research.
[4] Both authors contributed equally to this work.
[5] To whom correspondence should be addressed: Dept. of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium. Tel.: 32-3-265-23-23; Fax: 32-3-265-22-48; E-mail: sylvia.dewilde@uantwerpen.be.

[6] The abbreviations used are: 7TM, 7-transmembrane domain; GD, globin domain; RR, resonance raman; ESE, electron spin echo; PDB, Protein Data Bank; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; r.m.s.d., root mean square deviation; swMb, sperm whale Mb; β-AR, β-adrenergic.

pled receptor domain. In these receptors, ligand binding causes a conformational change, allowing a guanine nucleotide exchange in a G-protein. Their activating ligands vary widely in structure and characteristics (11–13) and include neuropeptides (14–16). In *C. elegans*, more than 100 neuropeptides have been identified by bioinformatic and mass spectrometric methods (17–19). They can be divided into three families as follows: the FMRF-amide (Phe-Met-Arg-Phe-NH_2)-related peptides or FLPs, the insulin-like peptides, also called INSs, and the remaining neuropeptide-like, non-insulin, non-FLP peptides, the NLPs (14, 15, 17, 20–22). Although a specific functional role has been identified for only few of these neuropeptides, overall they have been shown to play a crucial role in behavioral aspects as diverse as appetite (23, 24), fat storage (25), locomotion (26–30), reproduction (31), and dauer formation (32). Peptides exert their biological functions via specific signal-transducing receptors by direct or indirect modulation of the synaptic activity or by acting as primary neurotransmitters.

The C-terminal domain of GLB-33 (residues 373–534) displays the major determinants of the globin fold, including eight α-helices (A–H), the proximal His(F8) and the characteristic pattern of hydrophobic/hydrophilic residues (33, 34), and is therefore named as GLB-33 globin domain (GLB-33 GD). Despite the occurrence of specific deviations from standard templates like Val(CD1) and Ile(E7), this domain can be considered as a genuine globin (3).

To understand the functional role of the GLB-33 neuronal receptor, we have expressed and purified the recombinant GLB-33 GD in *Escherichia coli* and investigated its spectroscopic characteristics, kinetics, and equilibrium of the binding of the gaseous ligands O_2 and CO, and the reactivity with nitrite. We also performed *in vitro* and *in vivo* localization studies. Furthermore, models of the three-dimensional structures of the globin domain, as well as of the G-protein-coupled receptor domain, were built and used as templates to search the databases for proteins with similar structures. We have identified a chimeric protein in the nematode *Ascaris suum* with a similar transmembrane domain but with a structurally different globin domain. From this investigation, we have identified several structural and functional characteristics of the globin domain of GLB-33 that are likely to be involved in the fine-tuning of the neuronal receptor response in *C. elegans*.

EXPERIMENTAL PROCEDURES

In Vitro Localization—The cDNA coding for GLB-33 (*glb-33*) is 1629 bp long and was cloned in the pEGFP-N1-vector with C-terminal GFP tag (Clontech) using BglII and BamHI restriction enzymes (Biolabs, Westburg, Netherlands) and T4 DNA ligase (Novagen). 1 μg of the expression construct was transfected into human neuroblastoma SH-SY5Y cells (ATCC® CRL-2266) with 3 μl of Lipofectamine 2000 (Invitrogen). To obtain detailed images of the subcellular localization of GLB-33, an UltraVIEW Vox ERS microscope (PerkinElmer Life Sciences) was used, and images were created with the Volocity 6.0.1 software. GFP was excited at 488 nm with an argon laser, and a bandpass filter was used to allow emission light between 500 and 550 nm.

In Vivo Localization—The translational reporter for *glb-33* was constructed using fusion PCR, as described by Hobert (35). The reporter contained the *glb-33* target gene and 2.95 kb upstream and 0.653 kb downstream of the target gene to include the endogenous promoter region and 3′UTR gene regulating elements, respectively. The *gfp* gene was amplified from the pPD95.75 vector and fused at the 3′ side of the *glb-33* gene, thereby preceding the 3′UTR region. A coding region for a GA linker between GLB-33 and GFP was also included. The sequences of the primers used were as follows: 5′-TAGACAGTTTGCCACATCTT-3′ (forward primer of the promoter of *glb-33*); 5′-GGGTACACTCTTCAAAAACA-3′ (nested forward primer of the promoter of *glb-33*); 5′-TCCAGCGCCTGCACCAGCTCCTGGCATTTGACCAATTGTG-3′ (reverse primer of the *glb-33* gene); 5′-GCATGGATGAACTATACAAATGACTTTTTCCTAGATCTCGC-3′ (forward primer of the 3′UTR region of *glb-33*); 5′-TGTTCTTCTCCCGATATGGTAND-3′ (reverse primer of the 3′UTR region of *glb-33*); 5′-GACTCCGTGACAATTTCCTC-3′ (nested reverse primer of the 3′UTR region of *glb-33*); and 5′-GGAGCTGGTGCAGGCGCTGGAGCCGGTGCCTTGAGTAAAGGAGAAGAAC-3′ (forward primer of *gfp*), and 5′-TTTGTATAGTTCATCCATGCC-3′ (reverse primer of *gfp*).

In addition, the *unc-119* gene, including a 2.189-kb upstream region and a 1.228-kb downstream region of the *unc-119a* isoform, was amplified with forward primer 5′-TCAGTAAAAGAAGTAGAAT-3′ and reverse primer 5′-GAATTTTAACAATACTTC-3′. The PCR product was used as a co-injection marker and rescued the locomotion defect of the *unc-119*(*ed3*) strain, the strain used for micro-injection. The final PCR products were injected into the gonads of young adult hermaphrodites using an Axiovert 135 (Zeiss) microscope and FemtoJet microinjection system (Eppendorf), at a concentration of 50 ng/μl for the *glb-33* reporter and 20 ng/μl for the *unc-119* gene. Transformed lines were analyzed using a Nikon Eclipse TE2000–5 confocal microscope.

In Silico Identification of Orthologs—The protein sequence of GLB-33 (accession number AAK68603) was used to search the nonredundant database of NCBI for orthologs using the protein-protein Basic Local Alignment Tool (BLAST) algorithm (BLASTP) (36, 37). Additional similarity searches were performed on the *A. suum* CDS gene set (38) using the NemaBLAST algorithm available at Nematode.net version 3.0 (39). GLB-33 orthologs were also identified using the BLASTP algorithm implemented in WormBase (40). All significant hits were manually aligned with GLB-33 using GeneDoc (41), employing the procedure used earlier (42, 43) based on the myoglobin fold (33) and the pattern of predominantly hydrophobic residues at 37 conserved solvent-inaccessible positions.

Cloning, Expression, and Purification of Recombinant GLB-33 GD and A. suum FMRF GD—*C. elegans* worms were grown as described previously (2). Young adult worms were collected, and total RNA and cDNA were prepared as described (6). To confirm the presence of a globin domain in the *A. suum* FMRF-amide receptor (accession number ADY45777.1), the cDNA coding for this globin domain was amplified and sequenced. *A. suum* cDNA (1365 bp) was kindly provided by Peter Geldhof and Edwin Claerebout (University of Ghent, Belgium) and was prepared as follows. Total RNA from L3 larvae

was extracted with TRIzol reagent. After treatment with DNase (Invitrogen), cDNA was generated from 1 μg of total RNA using the cDNA synthesis kit (iScript™, Bio-Rad). PCR was then performed with the gene-specific forward and reverse primers 5′-ATGTGGCGTCAAAAACGGCCAC-3′ and 5′-TGGTTTCTTAAGGAAACTCTGTCC-3′, respectively. The PCR was performed in a final volume of 25 μl under the following conditions: 1× buffer, 1.5 mM $MgCl_2$ (Invitrogen), 0.2 mM of each dNTP, 1 μM of each primer, and 1 unit of Platinum® TaqPCRx DNA polymerase (Invitrogen). The PCR was started at 4 min at 95 °C, which was followed by 35 cycles consisting of 30 s at 95 °C, 30 s at 55 °C, and 60 s at 72 °C. A final step of 10 min at 72 °C was added to the last cycle. The PCR product was analyzed on a 1.5% agarose gel.

The cDNAs coding for the GLB-33 GD (*glb-33 gd*, bp 1120 to 1629) and for the *A. suum* FMRF GD (*A. suum fmrf gd* bp 991 to 1365) were cloned in the pET23a vector with C-terminal His tag (Novagen) using NdeI and XhoI restriction enzymes (Biolabs). The sequences of the forward and the reverse primers that were used to amplify the cDNAs were as follows: 5′-CATATGCTTCTCGGAGACAGGTTGTCC-3′ and 5′-CCGCTCGAGTGGCATTTGACCAATTGTGTC-3′ for *glb-33 gd* and 5′-GGAATTCCATATGTGGCGTCAAAAACGGCCAC-3′ and 5′- CCGCTCGAGTGGTTTCTTAAGGAAACTC-3′ for *A. suum fmrf gd,* respectively. Ligation of the cDNAs in the pET23a vector was performed using T4 DNA ligase (Novagen). The ligated product was sequenced on both strands using BigDye terminator chemistry and an ABI 377 sequencer (Applied Biosystems). *In vitro* expression in BL21(DE_3)pLysS *E. coli* cells was performed as described previously (44). Alternatively, cells were resuspended in 50 mM Tris, pH 7.5, and 300 mM NaCl. GLB-33 GD and *A. suum* FMRF GD were purified by nickel affinity chromatography (Clontech), eluted with 250 μM imidazole in resuspension buffer, and dialyzed against 20 mM Tris, pH 8.5, and 50 mM NaCl.

Optical, Resonance Raman, and Electron Paramagnetic Resonance Spectroscopy—UV-visible spectra were measured in a 250–700-nm range on a Cary-5 UV-visible NIR spectrophotometer (Varian). Expressed GLB-33 GD was dialyzed against different buffers with pH values ranging from 4 to 8.5 (100 mM sodium acetate, pH 4, 100 mM sodium acetate, pH 5, 100 mM BisTris, pH 6, 100 mM HEPES, pH 7, 100 mM Tris, pH 8.5). Ferrous CO-bound and reduced deoxy protein samples were prepared by flushing 1 ml of 100 mM potassium phosphate buffer, pH 7.0, for 15 min with CO and N_2 gas, respectively, in a sealed cuvette. After addition of 10 μl of a saturated solution of sodium dithionite, highly concentrated recombinant purified protein was added to obtain a final concentration of 50 μM. The ferrous oxy derivative of GLB-33 GD was prepared by reducing the protein with 10 μl of a saturated sodium dithionite solution, prior to loading the sample onto a PD10 gel filtration column (Amersham Biosciences) to remove the dithionite and allow O_2 to bind.

Resonance Raman (RR) measurements were carried out on an 80-cm Dilor XY-800 Raman scattering spectrometer consisting of a triple spectrograph operating in normal mode and a liquid nitrogen-cooled CCD detector. The excitation source was a Kr ion laser (Spectra Physics 2020) at 413.1 nm. The protein solution was stirred at 6000 rpm to avoid local heating. Five spectra (120-s recording time) were acquired and averaged after the removal of cosmic ray spikes by a program developed in-house. Laser powers of 1 and 50 milliwatts were used.

X band electron paramagnetic resonance (EPR) measurements were performed on a Bruker EleXsys instrument equipped with a helium cryostat (Oxford Inc.). All measurements were done at 10 K. The continuous wave EPR spectra were obtained with a modulation frequency of 100 kHz and a modulation amplitude of 0.5 millitesla. The electron spin echo (ESE)-detected EPR experiments were performed using the $\pi/2$-τ- π-τ-echo sequence with $t_{\pi/2}$ = 16 ns, t_{π} = 32 ns; τ was varied in steps of 8 ns from 88 to 3280 ns. The spectra recorded at the different τ values were subsequently summed.

CO Association Kinetics by Flash Photolysis—CO-bound GLB-33 GD samples were prepared in a sealed 4 × 10-mm quartz cuvette containing 1 ml of 100 mM potassium phosphate buffer and 1 mM EDTA at pH 7.0. The buffer was equilibrated with different mixtures of CO and N_2 gas, using a High-Tech system (Bronkhorst, The Netherlands). To obtain the association rate constant for CO binding ($k_{on,\,CO}$), CO-ligated ferrous GLB-33 GD was prepared with CO concentrations ranging between 200 and 800 μM. 10 μl of a saturated solution of sodium dithionite was then added to the buffer to scavenge any residual O_2. A minimal amount of concentrated protein was injected to obtain a final concentration of ~5 μM. The heme iron atom was reduced by the added dithionite, and conversion to the CO-ligated form was verified by UV-visible spectroscopy.

Flash photolysis experiments were carried out on a laser photolysis system (Edinburgh Instruments LP920) at 20 °C, using a frequency-doubled Q-switched Nd:YAG laser (Spectra Physics Quanta-ray) at 532 nm. CO association was measured by photo-dissociating ligated GLB-33 GD by a short laser pulse (5–8 ns) and then following the recombination of the photo-dissociated CO-ligand at different time scales, ranging from 2000 ns for the geminate recombination of CO to 400 ms for the bimolecular rebinding of the ligand at 421 nm.

Exponential decays from the consecutive time scales were joined together to give the complete ligand-rebinding curve. All data were analyzed using the Matlab program (The Math Works Inc.). The rate of geminate rebinding (k_{gem}) was obtained by fitting a single exponential curve through the data points collected in the first 2000 ns after photodissociation shown in Equation 1,

$$\Delta OD_t = \Delta OD_{gem} \cdot \exp(-k_{gem} t) + \Delta OD_{obs} \qquad \text{(Eq. 1)}$$

After logarithmic resampling of the data points, pseudo-first order fast observed rebinding rate constants ($k_{obs,\,f}$) and slowly observed rebinding rate constants ($k_{obs,\,s}$) were determined by least square fitting of the decays with bi-exponential Equation 2,

$$\Delta OD_{obs} = \Delta OD_f \cdot \exp(-k_{obs,f} \cdot t) + \Delta OD_s \cdot \exp(-k_{obs,s} \cdot t) \qquad \text{(Eq. 2)}$$

Fast and slow CO-rebinding association rate constants ($k_{on,\,f}$ and $k_{on,\,s}$, respectively) were calculated from the dependence of $k_{obs,\,f}$ and $k_{obs,\,s}$ on the CO concentration according to Equations 3 and 4 (45),

$$k_{obs,f} = k_{on,f} \times [CO] + k_{off,f} \quad \text{(Eq. 3)}$$

$$k_{obs,s} = k_{on,s} \times [CO] + k_{off,s} \quad \text{(Eq. 4)}$$

O_2 Equilibria—O_2 equilibrium curves of GLB-33 GD were determined at 25 °C using a modified thin layer diffusion chamber technique described elsewhere (46–48). Prior to measurements, the sample was reduced with dithionite and desalted, as described above. A 4-μl sample (~100 μM GLB-33 GD in 100 mM sodium phosphate, 0.5 mM EDTA, pH 7.4) was equilibrated with water-saturated gas mixtures of varying O_2 tension (PO_2) created by Wösthoff gas mixing pumps. Reversible absorbance changes following changes in PO_2 were recorded at 436 nm. Values for P_{50} (PO_2 at half-saturation) and n_{50} (cooperativity coefficient) were calculated from the zero intercept and slope of Hill plots, $\log(Y)/(1 - Y)$ *versus* $\log PO_2$, respectively, where Y is the fractional saturation.

Nitrite Reductase Activity—To study the reaction between deoxygenated GLB-33 GD and nitrite, a stock solution of nitrite (~20 mM in 100 mM sodium phosphate, 0.5 mM EDTA, pH 7.4) and a sample containing GLB-33 GD (~10 μM heme in 100 mM sodium phosphate, 0.5 mM EDTA, pH 7.4) were deoxygenated by equilibrating with N_2 gas for ~30 min. Nitrite solutions were freshly made every day, and nitrite concentrations were checked by the Griess method (49). An ~2–3-fold molar excess of a previously degassed dithionite solution was added (~20–30 μM final concentration) to the N_2-equilibrated GLB-33 GD sample, and absorbance spectra were taken to verify full heme deoxygenation. The reaction between deoxy-GLB-33 GD and nitrite (range 0.1–0.3 mM) was measured at 25 °C under pseudo-first order conditions in the presence of dithionite, as described previously (50, 51). Observed rates were obtained by fitting absorbance traces at 579 nm to a single exponential function, and the apparent second order rate constant was obtained from the slope of a linear plot of observed rates as a function of nitrite concentration.

Autoxidation Rate—The autoxidation rate of GLB-33 GD (~10 μM heme) was measured in air in 100 mM sodium acetate buffer, pH 4, and in 100 mM sodium phosphate, 0.5 mM EDTA, pH 7.4, at 25 and 37 °C, after previous reduction of the ferric protein with dithionite and desalting.

The decrease in absorbance at 577 nm was monitored spectrophotometrically every minute for the first 90 min and every 5 min for the next 60 min, using a Cary-5 UV-visible NIR spectrophotometer (Varian) in experiments performed at pH 4.0, and every 50 min using an HP 8543 UV-visible diode array spectrophotometer in experiments performed at pH 7.4. Absorbance traces were fitted according to a single exponential function.

Three-dimensional Modeling of GLB-33 and A. suum FMRF-amide Receptors—The three-dimensional structures of GLB-33 GD (residue 331–455), GLB-33 TM (residue 1–373), *A. suum* FMRF GD (residue 331–455), and *A. suum* FMRF 7TM (residue 1–330) were predicted using the I-TASSER server (52, 53) without structural restraints. The I-TASSER server builds three-dimensional models using multiple-threading alignments by LOMETS and iterative TASSER assembly simulations. The Phyre2 program was used to identify proteins with a similar fold, as it uses the alignment of hidden Markov models via HHsearch to significantly improve the accuracy and the detection rate (54). The quality of the models was evaluated with Procheck (55) and Verify3D (56). ProSA-web (57) was utilized to examine the energy of residue-residue interaction using a distance-based pair potential approach in which the energy is transformed to a Z score where residues with a negative score indicate reasonable side chain interactions. The overlays between the predicted models and the figures were made using the PyMOL Molecular Graphics System (58). The full-length predictions were generated with a combination of manual approach and docking techniques. First, the modeled N-terminal segment of the globin domains was approached to the C-terminal segment of the respective modeled 7TM domains. Second, the RosettaDock program (59) was launched to optimize the interface contacts of the two domains. A dimer of two GLB-33 GD models was also predicted using the known dimeric GLB-1 crystal structure (PDB code 2WTH) as a template. The interface of the dimer was optimized with the RosettaDock program.

RESULTS

In Silico Search for Orthologs of GLB-33—GLB-33 orthologs were found in the cluster of the Rhabditoidea (a subgroup of the Rhabditina, Nematoda) (60), more specifically in the genomes of *Caenorhabditis briggsae*, *Caenorhabditis remanei*, *Caenorhabditis brenneri*, *Caenorhabditis japonica*, *Caenorhabditis angaria*, *Caenorhabditis* sp5 and *Caenorhabditis* sp11, which are all closely related species belonging to the *Caenorhabditis* genus as well as in *Heterorhabditis bacteriophora*. Furthermore, by computational analysis we identified a partially annotated orthologous sequence of GLB-33 in the recently sequenced genome of *A. suum* (gene set CDS database, GS_17715). Despite conservation of two out of three intron positions (EF10.0 and FG7.0), the predicted gene only encoded a truncated form of the C-terminal globin domain and no N-terminal transmembrane domain. Most intriguingly, our comprehensive *in silico* searches also revealed similarities to a predicted FMRF-amide receptor in the *A. suum* genome (*A. suum* FMRF, GS_06644), with an *E*-value of $2e^{-15}$ and a sequence identity of 24% compared with GLB-33. To exclude whether no orthologs were missed, we screened the sequenced genomes of all nematodes, including *Ancylostoma* with GLB-33 GD, GLB-33 7TM, GLB-33, *A. suum* FMRF GD, *A. suum* FMRF 7TM, and *A. suum* FMRF as a template. No other receptors with a C-terminal extension similar to GLB-33 and *A. suum* FMRF were found.

Sequence Analysis of A. suum FMRF GD—Alignment of the protein sequences of GLB-33 and *A. suum* FMRF (24% identical and 44% similar) showed that the physicochemical properties of the majority of the amino acids are conserved (Fig. 1). In the alignment, the positions of the transmembrane helices and α-helices are indicated as predicted by the three-dimensional model of these proteins. It is clear that the amino acids building up the transmembrane helices are very similar, differing only by a few amino acids, thus resulting in transmembrane regions of comparable length (Fig. 1*A*). However, in the globin-like domain of *A. suum* FMRF, the lack of the A- and D-helices and

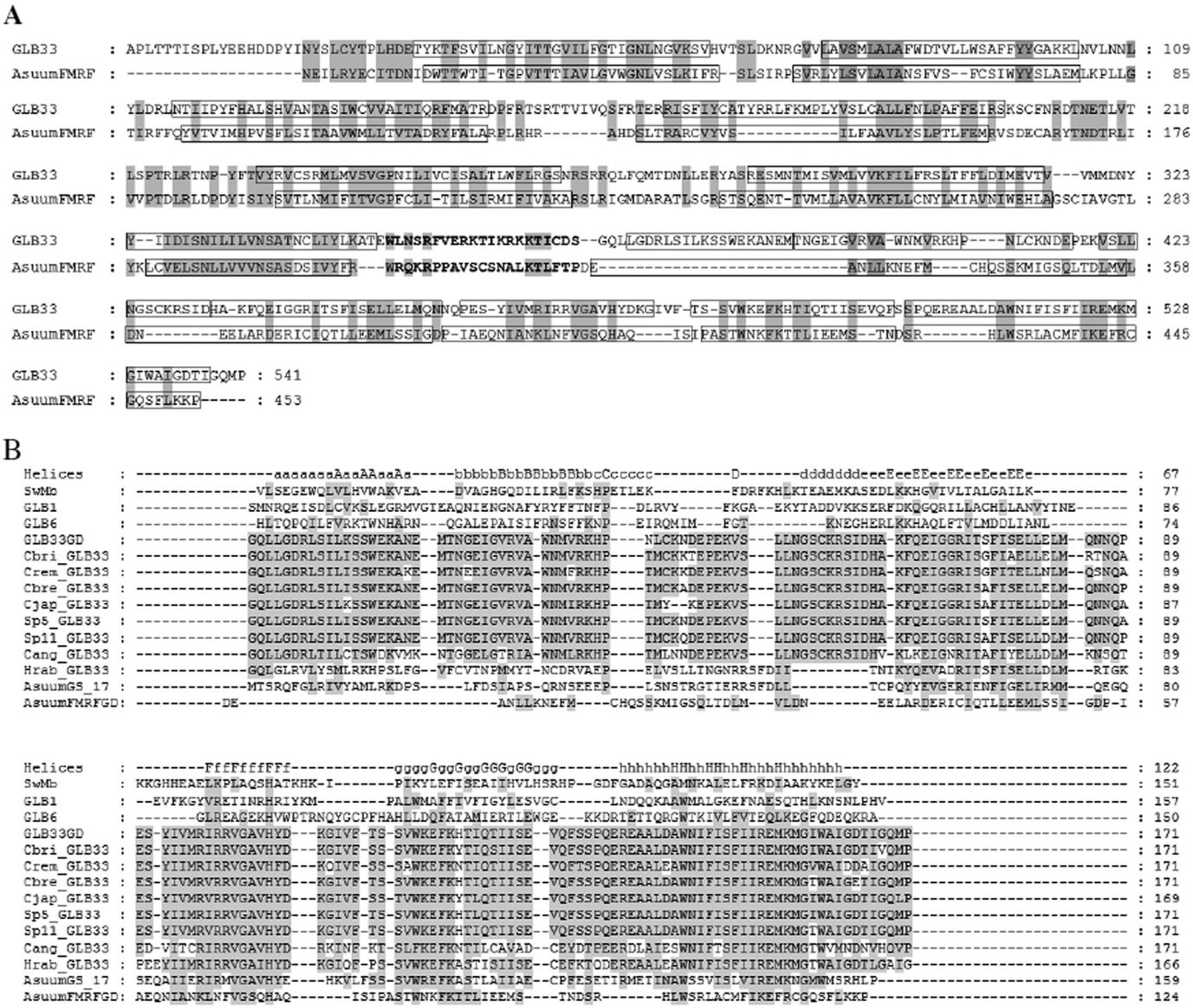

FIGURE 1. **Alignments based on the three-dimensional structures and three-dimensional models of the 7TM and globin domains.** *A*, GLB-33 and *A. suum* FMRF. Transmembrane helices and α-helices are *framed*. *Grayscale shading* indicates iso-functional or identical amino acid residues. The linker region is depicted in *bold*. *B*, GLB-33 with the relevant reference proteins swMb (accession number BAF03579), GLB-1 (accession number CAA77458), GLB-6 (accession number CCD65109), and orthologs in *C. briggsae* (Cbri, WormBase ID CBG06424), *C. remanei* (Crem, WormBase ID CRE05562), *C. brenneri* (Cbre WormBase ID CN07673), *C. japonica* (Cjap, WormBase ID JA64867), *C. angaria* (Cang, WormBase ID Cang_2012_03_13_00952.g15095.t1), *C.* sp5 (WormBase ID Csp5_scaffold_00786.g15463.t1), *C.* sp11 (WormBase ID Csp11.Scaffold506.g2439.t1), *H. bacteriophora* (Hrab, WormBase ID Hba_01784). and *Ascaris* (WormBase ID GS_17715, GS_06644). Helices are the annotated helices from the crystal structure of swMb. Conserved residues at buried sites in the *Helices* are indicated with a capital letter according to Ref. 33. *Grayscale shading* indicates iso-functional or identical amino acid residues.

the extended C-helix is striking when aligned to GLB-33 GD (Fig. 1, *A* and *B*), resulting in a possible modified truncated globin domain. This globin domain is preceded by a helical linker region that connects it to the 7TM domain (Fig. 1*A*). Therefore, we considered the possibility that the cDNA sequence of the globin domain of the *A. suum* FMRF reported in the database was incorrectly predicted, requiring a correction of the open reading frame. We therefore amplified the cDNA of *A. suum* FMRF GD and determined its sequence. Remarkably, the amplified cDNA was identical to the predicted one, supporting a correct protein sequence with an unusual tertiary structure.

In Vivo and in Vitro Localization of GLB-33—For the localization of GLB-33 *in vivo*, a translational reporter for the *glb-33* gene was constructed including the endogenous promoter region and 3′UTR elements. The expression pattern of this reporter was in line with the results of a previously used transcriptional reporter (4) and indicated that GLB-33 was exclusively expressed in the nervous system (Fig. 2*A*). The GLB-33 protein was present in a large number of neurons in the head and tail region, the nerve ring, the ventral and dorsal nerve cord, and several lateral nerve cords (Fig. 2, *B* and *D*). However, despite its wide expression in the nervous system, GLB-33 did not seem to be expressed in any of the amphid, cephalic, labial, or phasmid sensory neurons; the typical expression pattern for these neurons was not observed nor was any overlap seen between GFP expression and 1,1′-dioctadecyl-3,3,3′,3′-tetramethylindocarbocyanine perchlorate (DiI) staining, a

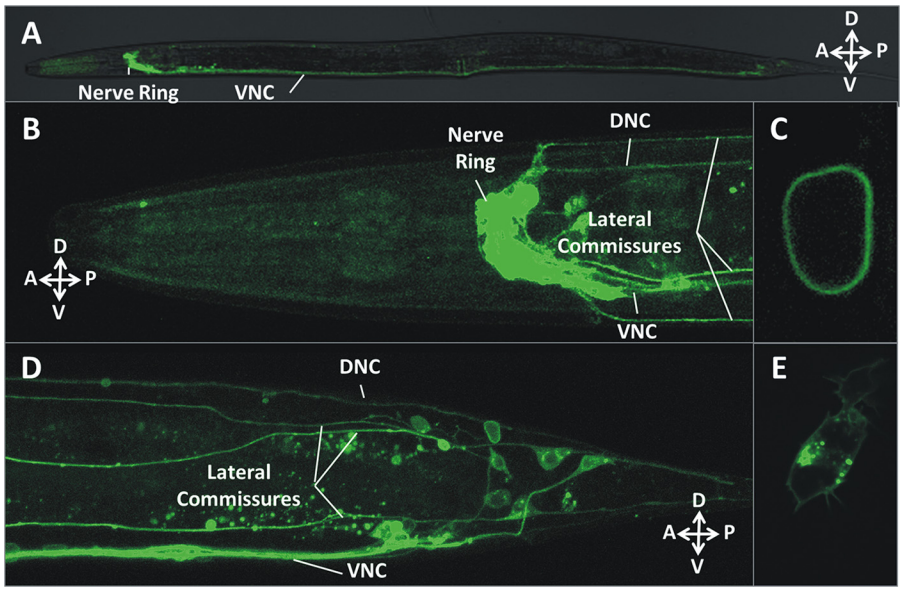

FIGURE 2. ***In vivo* and *in vitro* expression pattern.** *A,* overview of the *in vivo* expression pattern of the *glb-33* translational reporter. *B,* head region. *C,* subcellular expression of the reporter; *D,* tail region of the worm. GLB-33 is membrane-bound and is expressed in a high number of neurons in the head and tail region of the worm, the nerve ring, the ventral (*VNC*) and dorsal nerve cord (*DNC*), and several lateral nerve cords. *E, in vitro* membrane localization of GLB-33 in human neuroblastoma SH-SY5Y cells.

compound that selectively stained several amphid and phasmid neurons (data not shown). This indicates that GLB-33 is expressed in motor neurons or interneurons, which are involved in locomotor behavior or information processing, and not in neurons that sense environmental cues. Finally, the expression pattern of the reporter was in line with the membrane-bound localization of the receptor (Fig. 2*C*). The membrane localization of GLB-33 was confirmed by the transfection of *glb-33-gfp* in human neuroblastoma SH-SY5Y cells (Fig. 2*E*).

Spectral Characterization of GLB-33 GD and A. suum FMRF GD

UV-visible Spectra—The UV-visible absorption spectra of GLB-33 GD in the as-expressed, the dithionite-reduced (ferrous deoxy), the ferrous CO-bound, and the ferrous oxy form, prepared at pH 8.5 and the aquo-met form, prepared at pH 4, are shown in Fig. 3. The ferrous deoxy form has the Soret band at 433 nm and a Q band maximum at 558 nm. These maxima are typical for pentacoordinate high spin ferrous deoxyheme proteins (45). The absorption spectrum of the CO form of GLB-33 GD is similar to those of sperm whale myoglobin (swMb), with peaks at 420, 539, and 568 nm (45). After reduction with dithionite and desalting, the ferrous oxy form has peaks at 413, 543, and 581 nm (Fig. 3*A*). Although the absorption spectrum of as-expressed GLB-33 GD (pH 6–8.5) is apparently very similar to that of a ferrous oxyglobin, with the Soret band located at 413 nm and the Q bands at 544 and 577 nm, these maxima are also similar to those of the ferric hydroxide-ligated form of swMb (61) and to those of low spin heme oxygenase at alkaline pH (62). EPR (see below) confirms that the ligated state of the as-expressed GLB-33 GD can be assigned to the hydroxide-ligated ferric form. Note that in the as-purified samples at pH 8.5, there was still a small fraction of ferrous oxy form present capable of reversible O_2 binding.

The UV-visible spectra of the protein remained practically unchanged during the purification process, after several freeze-thaw cycles and after prolonged exposures to air, indicating that the hydroxide-ligated ferric derivative is very stable. Dialysis of GLB-33 GD against 10 mM sodium acetate, pH 4, resulted in the aquo-met form (Fig. 3*B*), with absorption maxima comparable with those of hexacoordinate high spin heme oxygenase (62). This acidic transition was completely reversible, because redialysis to buffers with pH 6–8.5 reconverted GLB-33 GD to the ferric hydroxide-ligated form, thus confirming that the as-expressed protein was not in the ferrous oxy state (data not shown).

In Fig. 3*C*, the UV-visible absorption spectra of the as-expressed and deoxy-reduced *A. suum* FMRF GD are depicted as prepared in 50 mM Tris-HCl, pH 7.5. The as-expressed protein has the Soret band at 412 nm and Q bands at 534 nm indicating that the globin is in a hexacoordinate low spin ferric state. Upon reduction, the Soret band shifts to 424 nm with Q band maxima at 53 and 558 nm, indicative of ferrous hexacoordinate globins, in contrast with the ferrous pentacoordinate heme of the deoxy-GLB-33 GD. The absorption maximum at 661 nm, typical for free biliverdin (63), in both spectra indicates the presence of denatured protein as the intensity of the peaks remained unchanged after reduction of the heme iron atom (Fig. 3*B*). Indeed, *A. suum* FMRF GD was very unstable, and extensive precipitation occurred during the whole purification process, as well as during attempts to bind other ligands to the heme iron atom, as could be deduced from the disappearance of the Q band maxima (data not shown). Stable liganded species could not be produced for this reason.

Resonance Raman Spectra—Fig. 4*A* shows the high frequency region of the RR spectra of GLB-33 GD in the as-expressed, the ferrous, and the CO-bound ferrous forms in 50 mM Tris-HCl, pH 8.5. The high frequency region of the RR spectra of heme proteins contains a number of so-called "marker bands" that depend on the oxidation state, the coordination state, and the spin state of the heme iron atom.

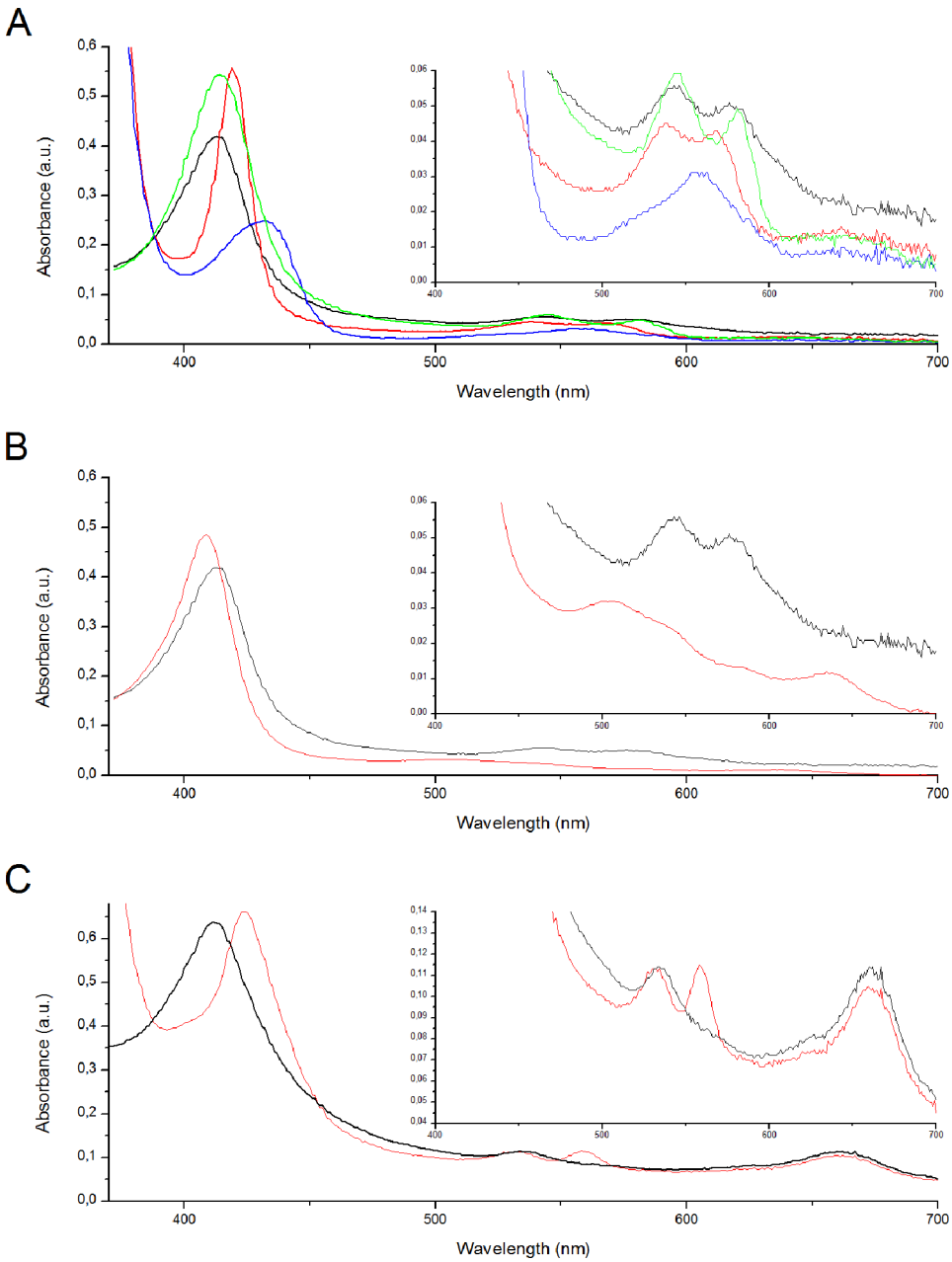

FIGURE 3. UV-visible absorption spectra of GLB-33 GD (*A*) in the as-expressed form pH 8.5 (*black*) (Soret band, 413 nm; Q bands, 577 and 544), the ferrous deoxy form (*blue*) (Soret band, 433 nm; Q band, 558 nm), the CO-bound ferrous form (*red*) (Soret band, 420 nm; Q bands, 568 and 539 nm), and the ferrous oxy form (*green*) (Soret band, 413 nm; Q bands, 581 and 543 nm) at pH 8.5. *B*, GLB-33 in the as-expressed form at pH 8.5 (*black*) and pH 4 (*red*). The latter gives a spectrum for the aquo-met form (Soret band, 408 nm; Q bands, 503 and 633 nm). *C*, *A. suum* FMRF GD in the as-expressed form (*black*) (Soret band, 412 nm; Q bands, 565 and 534 nm) and the deoxy-ferrous form (*red*) (Soret band, 424 nm; Q bands, 558 and 534 nm).

RR spectra confirmed that GLB-33 GD is expressed in the hydroxide-ligated ferric form under the conditions used (Fig. 4*A* (*a*)), with a first set of marker bands at $\nu_4 = 1376\ cm^{-1}$, $\nu_3 = 1502\ cm^{-1}$, and $\nu_2 = 1581\ cm^{-1}$. The ν_4 line at 1376 cm^{-1} and the ν_3 line at 1503 cm^{-1} are typical of hexacoordinate low spin ferric heme complexes (64). The frequencies of ν_3 and ν_2 are very close to the hydroxide-ligated form of met-globins, which are characterized with ν_3 at ~1505–1507 cm^{-1} and ν_2 at ~1585 cm^{-1} (65–67). A second set of marker bands is located at $\nu_3 = 1476\ cm^{-1}$ and $\nu_2 = 1558\ cm^{-1}$, which likely stems from a small fraction of a high spin ferric form (68, 69).

The high frequency region of dithionite-reduced, deoxy-GLB-33 GD (Fig. 4*A* (*b*)) has a single set of marker bands at $\nu_4 = 1356\ cm^{-1}$, $\nu_3 = 1470\ cm^{-1}$, and $\nu_2 = 1560\ cm^{-1}$, typical for a pentacoordinate high spin ferrous heme. Two ν_4 bands at 1371 and 1354 cm^{-1} are found in the RR spectra of the CO-bound ferrous protein recorded at low laser power (0.5 milliwatt) (Fig. 4*A* (*c*)). These stem from the CO-bound ferrous and photolysed form of the globin, respectively. The latter is confirmed by the fact that the intensity of the ν_4 band at 1354 cm^{-1} increases when the laser power is increased to 100 milliwatts (Fig. 4*A* (*d*)). This illustrates that the CO-bound ferrous form is photo-labile. The ν_4 frequency of the CO-bound form is lower than that of the as-expressed ferric form, suggesting that the GLB-33 GD distinguishes –OH from CO as a ligand, as also found for a bacterial heme-based aerotactic transducer (HemAT-*Bs*(O_2) (70)).

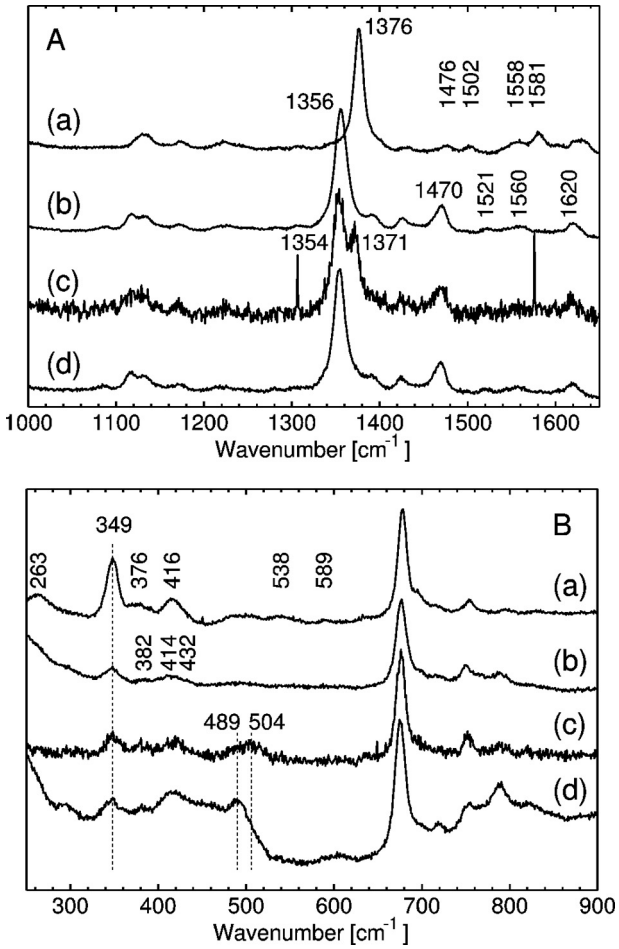

FIGURE 4. *A*, high frequency region, and *B*, low frequency region of the resonance Raman spectra of GLB-33 GD in the as-expressed hydroxide-ligated ferric form (*a*), the ferrous deoxy form at 50 milliwatt laser power (*b*), and the CO-bound ferrous form recorded at less than 0.5-milliwatt (*c*) and 100-milliwatt laser power (*d*).

Fig. 4*B* shows the low frequency region of the RR spectra of the different GLB-33 GD ligated forms. The low frequency region of the RR spectra of heme proteins contains a number of bending modes of the heme peripheral vinyl and propionate substituents. The frequency of these bending modes depends on the strength of the interactions between these side chains of the heme and amino acid residues in the heme pocket of the protein. The assignment of the RR bands is based on the work of Spiro and co-workers on Mb (68). The propionate bending mode, $\delta(C_{\beta C_c}C_d)$, of as-expressed GLB-33 GD (Fig. 4*B* (*a*)) is located at 376 cm^{-1} (broad signal) indicating a strong H-bond interaction between the heme propionates and nearby amino acid residues. A vinyl bending mode is found at 416 cm^{-1}. The broad profile of this band indicates the presence of a second vinyl mode at a higher frequency of ~430 cm^{-1}. A broader signal is found in the 490–515-cm^{-1} region where the $\nu_{Fe\text{-}OH}$ signal is expected (71). For dithionite-reduced deoxy-GLB-33 GD (Fig. 4*B* (*b*)), the frequency of the propionate bending mode is located at a higher frequency of 382 cm^{-1}, indicating an even stronger interaction between the heme propionates and nearby amino acids. Furthermore, a larger separation between the two vinyl modes was observed. One is located at 414 cm^{-1}, and the second mode is located at 432 cm^{-1}. These data show significant interactions between the protein moiety and the heme that are dependent on the heme-ligation state.

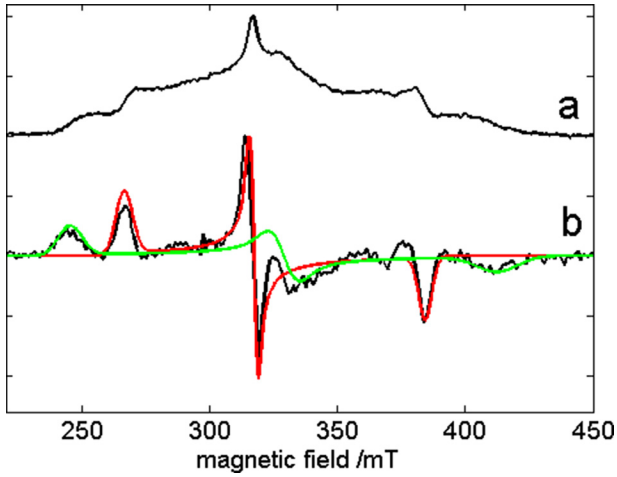

FIGURE 5. *a*, ESE-detected EPR spectrum of as-purified GLB-33 GD (sum of 400 τ values between 88 and 3280 ns); *b*, first derivative of spectrum *a* in *black* with simulation of ferric hydroxide GLB-33 GD (*red*) and ferric azide GLB-33 GD (*green*). The measurements were performed at 9 K.

The Fe-CO stretching mode can be found in the low frequency region of the RR spectra of CO-bound ferrous heme proteins (72). In the current case, a clear determination of the Fe-CO modes is hampered by the high photolability of the CO-ligated GLB-33 GD form (Fig. 4*B* (*c* and *d*)), although a broader signal is seen in the 489–504-cm^{-1} region typical of the Fe-CO modes.

EPR—The continuous wave EPR (data not shown) and ESE-detected EPR (Fig. 5*a*) both confirm the presence of two low spin ferric forms in the as-purified GLB-33 GD sample at pH 8.5. The two low spin components have the following principal g values: (g_1,g_2,g_3) = (2.62,2.20,1.815) (Fig. 5*b*, red) and (g_1,g_2,g_3) = (2.845,2.12,1.69) (Fig. 5*b*, *green*). The former g values are similar to those of a low spin Antarctic fish hemoglobin (73) and point to the formation of a hydroxide-ligated ferric heme form (74). The g values are also similar to those observed for alkaline cytochrome *c* peroxidases (75, 76). The second set of g values is very similar to that of horseradish peroxidases ((g_1,g_2,g_3)~(2.9,2.1,1.6)) showing low spin hydroxide ligation of the ferric heme in alkaline conditions (77, 78). A recent study on *Thermobifida fusca* hemoglobin revealed the presence of two ferric forms with similar EPR parameters as those found here. The component with a g_1 value around 2.8 was assigned to the hydroxide-ligated ferric complex where the OH group is strongly hydrogen bonded, whereas the form characterized by g_1 at ~2.6 is ascribed to a coordination mode in which the OH group is not so strongly H-bonded (79). Interestingly, also in *T. fusca* hemoglobin, the hydroxide-ligated complexes were still observable at pH 6.0. No significant contribution of a high spin ferric form (*i.e.* aquo-met) is found in the continuous wave EPR or ESE-detected EPR of GLB-33 GD at pH 8.5 (data not shown). The EPR spectra thus show that low spin ferric heme complexes are already dominantly formed at lower pH values than in most other globins. The UV-visible absorption and RR spectra are in agreement with this finding. The difference between the two low spin ferric forms most probably stems from a change in the hydroxide stabilization by the surrounding protein matrix.

TABLE 1

Rate constants of ligand binding of *C. elegans* globins compared with rate constants reported for other globins

Values for *GLB-26* and GLB-1 were taken from Ref. 6, for swMb from Ref. 101, and for *A. suum* Hb from Ref. 102. ND = not detectable.

	GLB-33 GD	GLB-1	GLB-26	swMb	*A. suum Hb*
CO					
$k_{on,f}$ (μM^{-1} s^{-1})	0.90 ± 0.10		23		
$k_{on,s}$ (μM^{-1} s^{-1})	0.35 ± 0.01			0.51 ± 0.06	0.21 ± 0.02
k_{gem} ($\times 10^7$) s^{-1}	9.87 ± 0.01			0.55 ± 0.08	3.0 ± 0.2
F_{gem}	0.50 ± 0.05			0.043 ± 0.006	0.45 ± 0.05
O_2					
k_{on} (μM^{-1} s^{-1})		23	ND	14	3
k_{off} (s^{-1})		0.5	ND	15	0.013
K_{O2} (μM^{-1})	≥8.55[a]	46[b]	ND	1.1	215
*P*50 (torr)	≤0.07[a]	0.003[b]	ND	1	0.0072
k_{oxid} (h^{-1}) pH 4	1.81 (25 °C) 15 (37 °C)				
k_{oxid} (h^{-1}) pH 7.4	0.09 ± 0.02 (25 °C)			0.054 (37 °C) 0.005 (25 °C)[c]	

[a] Data are from equilibrium experiments.
[b] Data are from kinetic experiments.
[c] pH 7.2, from Ref. 83.

Ligand Binding Kinetics of GLB-33 GD

CO Rebinding Kinetics—Flash photolysis was performed at pH 7 as described under "Experimental Procedures." CO rebinding from within the protein matrix to the heme iron atom is referred to as the [CO]-independent geminate rebinding phase (occurs in the nanosecond time scale after flash photolysis). For GLB-33 GD, this geminate rebinding process occurred with a rate constant of 9.87 (±0.07) × 10^7 s^{-1} (Table 1), about 18 times as fast as the geminate rebinding of CO to swMb (k_{gem} = 0.55 (± 0.08) × 10^7 s^{-1}, see Table 1). About half of the CO molecules rebind immediately, resulting in a high fraction of geminate rebinding of 0.50 ± 0.05 (Table 1), an indication that the CO molecules cannot easily escape from the protein matrix in comparison with swMb (F_{gem} = 0.043 ± 0.006, see Table 1). On longer time scales, a fraction of the gaseous ligands can escape to the bulk solvent and rebind from outside the protein matrix in a [CO]-dependent way. This bimolecular rebinding phase in GLB-33 GD is a biphasic process, characterized by a fast and a slow rebinding rate constant as follows: $k_{on,f}$ = 0.90 (± 0.10) × 10^6 M^{-1} s^{-1} and $k_{on,s}$ = 0.35 (± 0.01) × 10^6 M^{-1} s^{-1}, respectively (Table 1 and Fig. 6). The latter is comparable with the CO rebinding rate of swMb (Table 1).

O_2 Equilibria—The apparent P_{50} measured with the diffusion chamber method was 0.07 ± 0.01 torr at 25 °C, pH 7.4 (data not shown). However, we cannot exclude that at these very low levels of PO_2 the protein within the diffusion chamber was partially oxygenated even when equilibrated with pure nitrogen (*i.e.* to obtain zero saturation). Therefore, it is possible that the true P_{50} is lower than 0.07 torr. The equilibrium association constant (K_{O2}, M^{-1}) for O_2 derived from the P_{50} value is ≥8.5. μM^{-1} when assuming an O_2 solubility in water of 1.67 μM $torr^{-1}$ at 25 °C (80). A Hill coefficient close to 1 (~0.9) indicated absence of cooperativity (data not shown).

Autoxidation Rate of GLB-33 GD—Autoxidation of oxy-GLB-33 GD at pH 4 (25 and 37 °C) and at pH 7.4 (25 °C) showed a single exponential process with a rate of 1.81 h^{-1} (25 °C) and 15 h^{-1} (37 °C) at pH 4, ~278 times faster than native swMb (0.054 h^{-1}) at pH 7 and 37 °C (81), and ~62 times faster than the swMb H(E7)68E mutant (0.24 h^{-1}) at pH 7 and 37 °C (82). At pH 7.4 the rate constant is 0.09 ± 0.02 h^{-1}. As the absorbance peaks of the oxy- and hydroxide-ligated ferric forms are almost identical (Fig. 3*A*), the spectral change over time was small. However, the fact that isosbestic points (*e.g.* 531, 555, 568, and 585 nm at pH 7.4) were maintained over time (data not shown) is consistent with a single oxy to ferric transition at both pH 4.0 and 7.4. For pig myoglobin, a larger increase of the oxidation rate constant (several thousands) (82) was found when going from pH 7 to pH 4 in comparison with GLB33GD. However, this smaller increase can be due to the fact that there is a conversion from the hydroxide-ligated form to the aquo-met form for the latter.

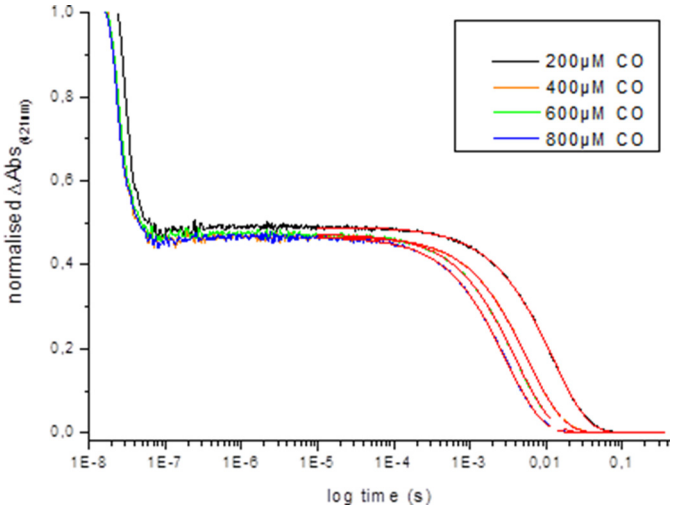

FIGURE 6. **CO rebinding to GLB-33 GD.** CO rebinding curves after flash photolysis recorded at 421 nm at different CO concentrations. *Continuous lines* in *red* are nonlinear least squares fitting of data obtained according to Equations 2 and 3.

Nitrite Reductase Activity of GLB-33 GD—The nitrite reductase activity was measured in the presence of a slight excess of dithionite. The data show that deoxygenated GLB-33 GD was able to convert nitrite to NO and the nitrite-induced conversion of deoxyheme to NO-bound heme was confirmed spectrophotometrically (Fig. 7*A*). The visible absorption spectrum changed over time from that of deoxy-GLB-33 GD (Fe^{2+}) with a single peak at 558 nm to that of a NO-bound globin (Fe^{2+}-NO) with peaks at 545 and 570 nm, indicating that the

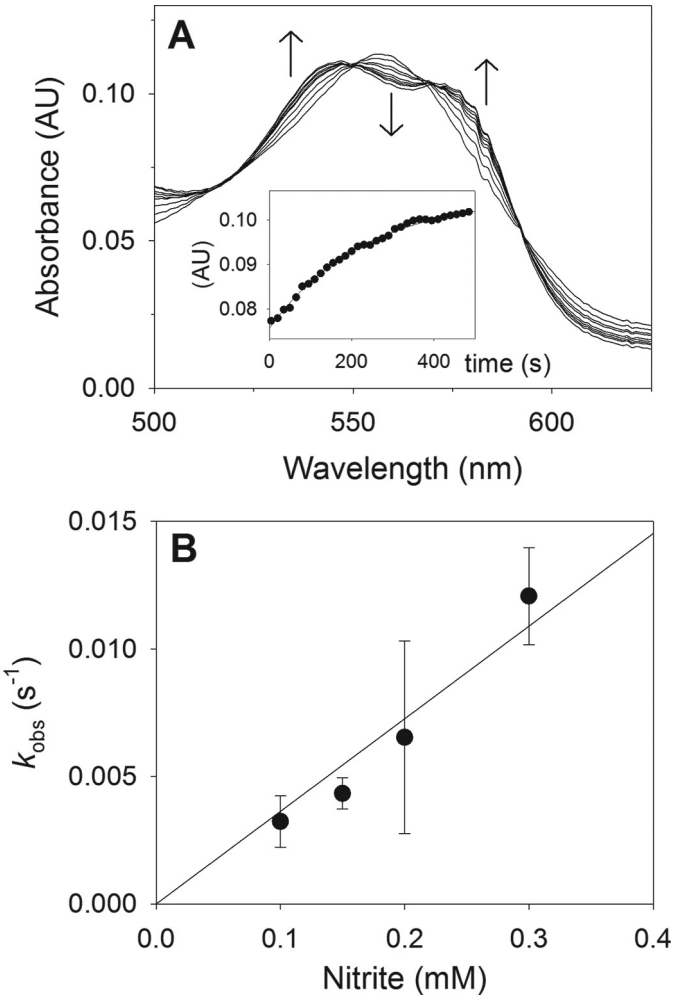

FIGURE 7. **Nitrite reductase activity of GLB-33 GD.** *A,* representative spectral changes accompanying the transition of deoxy-GLB-33 GD to NO-bound GLB-33 GD. *Arrows* indicate the direction of change. An example of monoexponential fitting of the absorbance trace at 579 nm over time (seconds) for the reaction between GLB-33 GD and 0.2 mM nitrite is shown in the *inset*. *B,* plot of k_{obs} as a function of nitrite concentration. The second-order rate constant is given by the slope of the linear regression (pH 7.4, 25 °C). *AU*, absorbance units.

reaction followed the general scheme for nitrite reduction by deoxyglobins in the presence of dithionite (51, 84). When plotting the fitted observed rates (k_{obs}) as a function of nitrite concentration, the second order rate constant was 36.3 M^{-1} s^{-1} (Fig. 7*B*), which is ~10-fold higher than that for Mb (2.9 M^{-1} s^{-1} (85).

Three-dimensional Modeling of GLB-33 GD and GLB-33 7TM Domain—The structural model of the N-terminal GLB-33 7TM domain was predicted using the I-TASSER software (52, 53). The best predicted model shows a C score of −0.30 and an expected TM score of 0.67 ± 0.12. Both these values indicate a satisfactory model with reliable topology. The stereo-chemical quality of the GLB-33 7TM model was analyzed with Procheck. The resulting Ramachandran plot displayed 91.9% of all amino acid side chain angles in allowed regions and 8.1% in disallowed regions. Verify3D analysis showed about 50% of the residues with a score over 0.15, and ProSA-web evaluation confirmed this result with an overall calculated Z score of −2.52. Phyre[2] identified the human β_2-adrenergic G-protein-coupled receptor (β_2-AR, PDB code 2RH1 (86)) as the most similar protein of known structure, with 41% similarity with the query sequence, and which was chosen as reference structure for the comparison. An overlay of the model with the β_2-AR structure was generated using PyMOL (Fig. 8*A*). A root mean square deviation (r.m.s.d.) value of 1.80 Å was calculated over 243 Cα atom pairs, suggesting the presence of several differences. The position of every helix of the modeled GLB-33 7TM is very well superimposed on the β_2-AR structure throughout the whole length, with exceptions in correspondence of numerous loops as follows: N- and C-terminal extremities as well as the loop connecting helix 5 to 6 are not present in the PDB file of β_2-AR (evidence of strongly flexible regions); the GLB-33 7TM model shows a 17-residue longer loop connecting helix 3 to 4 compared with the reference structure; and the amino acid stretch that connects helix 4 to 5 presents a short ordered helix in the β_2-AR structure, although it was predicted as disordered in our model. The most significant distances in the Cα positions of the α-helices are between a minimum of 1.8 Å (N-terminal of helix 5) and a maximum of 6.5 Å (N-terminal of helix 1).

Similarly, a prediction model was generated for GLB-33 GD. The C score (−0.62) and the expected TM score (0.63 ± 0.13) indicate the good quality of the best model predicted. Procheck evaluation indicated that the geometry of 97.6% of the residues is in allowed regions and as little as 2.4% is in disallowed regions. Verify3D showed that all the residues have a score higher than 0.2, and ProSA-web results confirm this finding through a negative Z score for every residue and a calculated Z score of −7.44 for the global molecule. All these data indicate a reliable model. For comparison, an overlay of the predicted model on the known high resolution three-dimensional structure of GLB-1 (PDB code 2WTG (6)) was made (Fig. 8*B*). The r.m.s.d. of the overlay resulted in 2.30 Å calculated on 134 Cα atom pairs. The 3-over-3 globin fold is conserved in the GLB-33 GD model; nevertheless, the different orientation of the modeled helices compared with the reference structure causes a global noncorrespondence of the Cα backbone. A closer look at the amino acid residues in the distal site of the GLB-33 GD model highlights its exceptional hydrophobic composition as follows: Leu-41(CD3), Ile-69(E7), Arg-72(E10), and Ile-73(E11) (Fig. 8*C*). This highly apolar environment leaves Arg-72(E10) as the only residue available for ligand stabilization. Likewise, *C. elegans* GLB-1 displays a majority of hydrophobic residues in its distal pocket (Phe-34, Phe-38, Phe-49, Phe-66, Ile-73, and Phe-114), but O_2 is directly stabilized by an unusual combination of the distal residues Tyr-35(B10) and Gln-69(E7) (6).

Until now, no three-dimensional structure of a globin associated with a transmembrane domain has been solved. As a consequence, homology modeling of the full-length protein was impossible. To visualize the full structure of GLB-33, a manual approach of the assembly of the GLB-33 GD model to the GLB-33 7TM model was attempted. The N terminus of the globin domain was located manually in the proximity of the C terminus of the 7TM domain. Subsequently, the docking program RosettaDock was used to optimize the interface interactions between the two models. The result is shown in Fig. 8*D*. Despite the limitation of this approach, the docking shows a

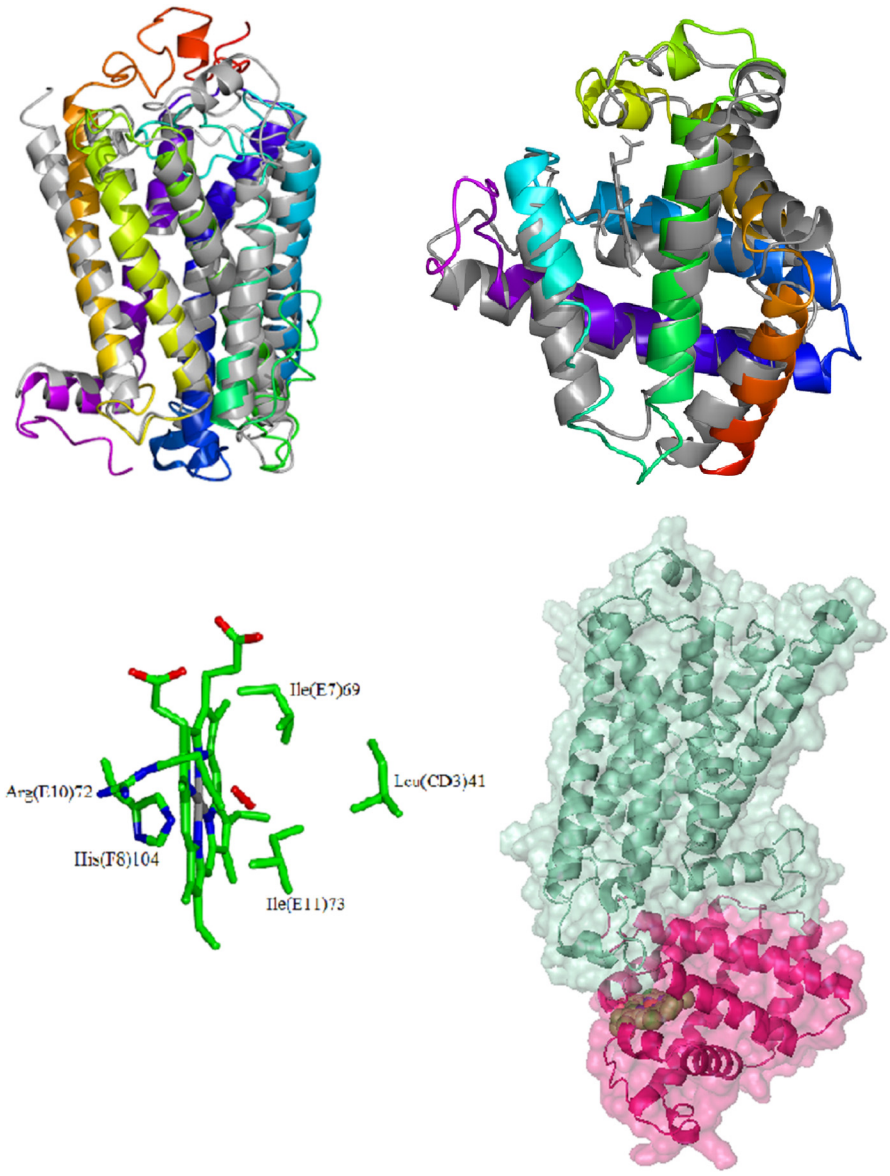

FIGURE 8. **Three-dimensional models of GLB-33.** *A,* three-dimensional model of GLB-33 7TM and overlay with the human β2 G-protein-coupled receptor. *B,* three-dimensional model of GLB-33 GD and overlay with GLB-1. *C,* detailed view of the heme pocket of GLB-33 GD. *D,* manual approach of the GLB-33-GD model to the GLB-33–7TM model using RosettaDock.

wide interaction surface between the two domains and a possible connection between the two terminal loops of the predicted models. Indeed, the interaction surface involves the N terminus of the A-helix, the E-F corner (*i.e.* C-terminal E-helix and N-terminal F-helix), 70% in length of the H-helix, and the C-terminal loop of the GD domain. Thus, these data suggest that any conformational change that occurs in the GD can be readily transmitted to the 7TM domain.

Three-dimensional modeling of A. suum FMRF 7TM and A. suum FMRF GD—Similarly to GLB-33, the full-length FMRFamide receptor of *A. suum* was modeled separating first the N-terminal *A. suum* FMRF 7TM from the C-terminal *A. suum* FMRF GD and second rejoining manually the two predicted domains. The *A. suum* FMRF 7TM model predicted by the I-TASSER software (52, 53) shows a C score and an expected TM score (0.41 and 0.77 ± 0.10, respectively) that indicates a good quality model. The geometry analysis performed with Procheck resulted in 94.5% of the residues in allowed regions and in 5.5% in disallowed regions. The Verify3D program highlighted that about 38% of the residues shows a score over 0.15 and the Z score calculated with ProSA-web of −0.49 confirmed this result. As for GLB-33–7TM, Phyre[2] identified the human β_2-AR (86) as the most similar protein of known structure with 35% similarity. An overlay of the *A. suum* FMRF 7TM model with the GLB-33 7TM model was generated using PyMOL (Fig. 9*A*). Over 282 Cα atoms, the r.m.s.d. of 1.18 Å was calculated, suggesting a low level of variety. Indeed, the two models superimpose very well, except for longer N- and C-terminal loops in the GLB-33 7TM model, and for the longer interhelical loop between helices 3 and 4 already noticed in its comparison with the β_2-AR structure, which is absent in the prediction of *A. suum* FMRF 7TM.

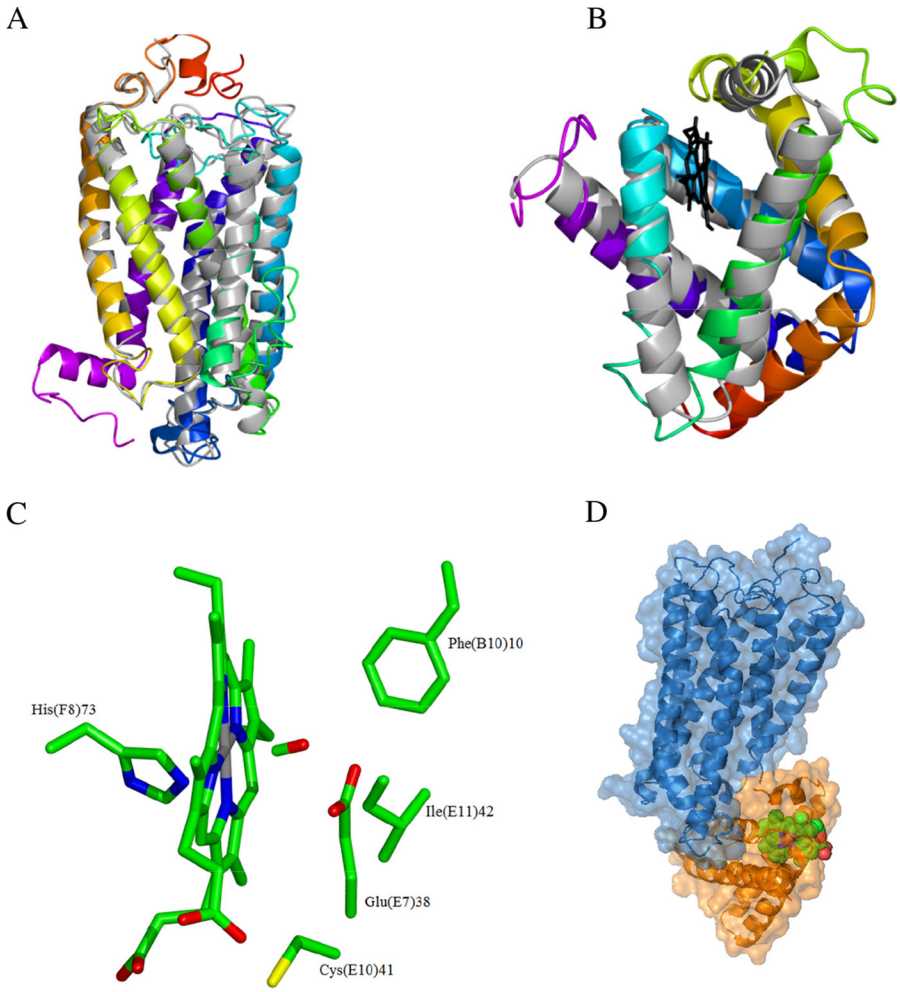

FIGURE 9. **Three-dimensional models of *A. suum* FMRF-amide receptor.** *A,* three-dimensional model of *A. suum* FMRF 7TM and overlay with GLB-33 7TM. *B,* three-dimensional model of *A. suum* FMRF GD and overlay with GLB-33 GD. *C,* detailed view of the heme pocket of *A. suum* FMRF GD. *D,* manual approach of the *A. suum* FMRF 7TM model and of the *A. suum* FMRF GD model using RosettaDocking.

Strikingly, the three-dimensional model for the putative C-terminal *A. suum* FMRF GD indicated a globin fold not referable to any known globin family. It can be recognized as a classical 3-over-3 α-helical globin fold, but lacking the A and D helices and exhibiting a longer C-helix (Fig. 9*B*). For this model, we obtained a C score of -0.90 and an expected TM score of 0.56 ± 0.14 with the I-TASSER software. The geometry quality assessment performed with Procheck indicated that 98.4% of the residues were in allowed regions and 1.6% were in disallowed regions. Verify3D and ProSA-web display positive concordant results (about 93% the residues have a score higher than 0.2, a negative Z score for every residue and a calculated Z score of -5.2 for the global molecule, respectively). All these data confirm the good geometry of this model. The Phyre[2] similarity search identified the globin domain of GLB-6 from *C. elegans* (PDB code 3MVC (7)) as the most similar structure with 42% of similarity. Because *A. suum* FMRF GD lacks several helices, we modeled its structure with a variety of 2/2 truncated hemoglobin structures as a template to exclude bias in the modeling program. These tests did not result in a good model however, and suggested that the α-helical fold of *A. suum* FMRF GD is more related to the 3/3 classical globin fold than to the 2/2 truncated hemoglobins, even though some deviations can be detected (data not shown). Superimposing the models of *A. suum* FMRF GD and of GLB-33 GD (Fig. 9*B*), we obtained the r.m.s.d. value of 2.57 Å over 88 Cα atom pairs. Obviously, this comparison emphasizes the lack of the A-, B-, and D-helices in *A. suum* FMRF GD. The shorter F-helix of GLB-33 GD is also noticeable, as well as the similar open conformation of the E-helix in both predicted models. A detailed look at the amino acid composition of the heme pocket reveals that it is more polar than the heme pocket of GLB-33 GD, with Phe-10(B10), Glu-38(E7), Cys-41(E10), and Ile-42(E11) (Fig. 9*C*).

Finally, the full-length appearance of *A. suum* FMRF was investigated using the model (Fig. 9*D*). The two docked domains display a discrete interaction area and the possibility of a covalent connection between the two termini of the predicted models. The heme pocket of the docked *A. suum* FMRF GD is even more exposed to the solvent than in GLB-33 GD (Fig. 9*D*), as shown in Fig. 9*D*. Obviously, this is also due to the peculiar truncated globin fold of *A. suum* FMRF GD.

DISCUSSION

The similarity in primary and tertiary structures between the *C. elegans* GLB-33 and the *A. suum* FMRF-amide receptor, as indicated by modeling analysis (Fig. 9), suggests that they could

serve a similar role *in vivo*. The modeled N-terminal extracellular loops create a structural pocket where FLPs could bind thus leading to a conformational change, resembling that of other FMRF-amide receptors (Fig. 8*A*). Nothing is known, however, about the process regulated by these chimeric proteins, but some indications can be found by studying the *in vivo* transcription pattern. Analysis of transcriptome data of *A. suum* during development indicates that the FMRF-amide receptor is predominantly expressed in the L3 larval stage, and peaks when the larvae have just hatched from the egg,[7] but it is unclear so far what role the FMRF-amide receptor could play in this developmental stage. In contrast, the *C. elegans* GLB-33 is expressed throughout all life stages of the nematode (3). Following the *in vivo* expression pattern of GLB-33, we detected this protein predominantly in inter- and motoneurons, and it might therefore be involved in movement following processing sensory cues (Fig. 2).

To explore the function of the globin domains of GLB-33 and of the *A. suum* FMRF-amide receptor, recombinant GLB-33 GD and *A. suum* FMRF GD were prepared and their biochemical properties investigated. It has to be noted that in *A. suum* the recombinant globin domain was very unstable, probably due to the lack of several helices as compared with genuine globins and the lack of the FMRF-amide receptor domain that might compensate this truncation. In the absence of various gaseous ligands, the globin domain of GLB33 showed an optical spectrum that is characteristic of a pentacoordinate ferrous globin, whereas the globin domain of *A. suum* is hexacoordinate. Raman and EPR spectra of *C. elegans* GLB-33 GD show tight interactions between the heme group or its ligand and the surrounding protein moiety (Figs. 4*B* and 5). Because changes in the ligation state of the heme iron atom are accompanied by conformational changes in the globin core, this suggests that the heme ligation and/or redox state of the globin domain could influence the conformational state of the coupled receptor domain. In this way, the globin domain could modulate the sensitivity of the effector domain for the binding of neuropeptides. Likewise, the binding of a neuropeptide to the receptor domain could introduce a conformational change in the globin domain, thus altering the ligand-binding properties of the latter. More arguments for the modulation of ligand affinity can be found in our modeling data. Docking the model of GLB-33 GD to that of GLB-33 7TM showed that the opening to the heme pocket is directed to the outer environment (Fig. 8*D*), suggesting that in the chimeric protein ligand binding and dissociation are still possible.

The reactivity of a globin domain toward ligands depends mostly on the amino acid residues in the heme pocket that can stabilize the bound ligand. The distal B10, E7, E10, and E11 residues are known to be involved in this stabilization in several globins (87–89) and are occupied by a different set of residues in the two globin domains investigated here (Fig. 1*B*): Ala-31, Ile-69, Arg-71, and Ile-73 in GLB-33 GD, and Phe-10, Glu-38, Cys-41, and Ile-42 in *A. suum* FMRF GD, respectively. Interestingly, neither globin domain contains the distal His(E7). Few examples of globins that contain Ile as the E7 distal residue, as GLB-33 GD, are known, including the truncated Hb of the cold adapted bacterium *Pseudoalteromonas haloplanktis* TAC125 (*Ph*-2/2HbO), where the B10, E7, E10, and E11 positions are occupied by Tyr, Ile, Val, and Phe, respectively (90). The electronic absorption spectra of *Ph*-2/2HbO at physiological pH and low temperature (12 K) are very similar to those of GLB-33 GD at the same pH at room temperature, with absorption maxima in the Q band region at 535 and 570 nm, indicating a hexacoordinate low spin ferric species with a hydroxide group as the 6th ligand. In the case of *Ph*-2/2HbO, this ligand is donated to the heme iron by the Tyr(B10) residue (90). At room temperature, the absorption spectrum of *Ph*-2/2HbO indicates the additional presence of a hexacoordinate high spin His-Fe-H_2O form (90) and is similar to the spectrum of GLB-33 GD at pH 4 (Fig. 3*B*). Also *T. fusca* hemoglobin, with Tyr, Ala, Arg, and Leu on positions B10, E7, E10, and E11, exhibits hydroxide-ligated forms, which are in equilibrium with an aquo-met form at pH 6 (79). In the latter protein, EPR contributions of hydroxide-ligated heme forms with similar principal *g* values as for the GLB-33 GD case were observed, with the exception that the g_1 ~2.6 contribution was only a minority species. In the case of GLB-33 GD, both components contribute approximately equal to the overall EPR spectrum. This is in line with the amino acid residues in the distal pocket of the two proteins. In *T. fusca* hemoglobin, the hydroxide group bound to the ferric heme is stabilized by Tyr(B10), Tyr(CD1), and Trp(G8) in agreement with the dominance of the g_1 ~2.8 contribution in the EPR spectrum indicative of strong H-bonding to OH^-. In contrast, in GLB-33 GD, the B10 position is occupied by an Ala residue unable to stabilize the hydroxide ligand. The only polar side chain in the largely apolar heme pocket is Arg(E10). This explains the larger fraction in which the heme-bound hydroxide is more weakly H-bonded to the protein than in the *T. fusca* hemoglobin, as shown by the larger EPR g_1 ~2.6 contribution (Fig. 5). Because no amino acid residues are present in the distal pocket of GLB-33 GD that can donate a hydroxide group to the ferric heme iron (Fig. 8*C*), it is most likely that the hydroxide ligand originates from a water molecule as is also the case for *T. fusca* hemoglobin. This hypothesis is supported by the observation that the conversion between the high spin aquo-met species and the low spin hydroxide-ligated form is completely reversible upon pH change through dialysis (data not shown).

Interestingly, a similar heme environment is found in the distal pocket of the Per-Arnt-Sim (PAS) domain of the O_2 sensor Fix-L of *Bradyrhizobium japonicum* (*B. japonicum* Fix-L). Although the structures of the globin domain and of the PAS domain differ significantly, they both bind a heme cofactor, and some comparisons can be made. In *Bradyrhizobium japonicum* Fix-L, ligand interaction is possible through the guanidinium group of $G_{\beta 2Arg}$ (91). This residue has been shown to assist in the sensing of O_2. This is not accomplished by stabilizing the oxy form however; instead, the side chain is involved in triggering a conformational switch upon binding of regulatory ligands (92). In GLB-33 GD, the Arg(E10) residue could be involved in a similar mechanism. It appears plausible that when O_2 binds with high affinity to the ferrous heme, it will induce a confor-

[7] Peter Geldhof (University of Ghent) and Robin Gasser (University of Melbourne), personal communication.

mational switch of the Arg-72(E10) residue upon formation of the oxy derivative, which may then act as a trigger to transduce the signal to the transmembrane receptor domain. The observation of two ferric hydroxide-ligated forms observed in EPR spectra (Fig. 5) suggests the existence of two different protein conformations with potentially different roles in signal activation or deactivation.

Another functionally important property of ferrous GLB-33 GD is its higher reactivity with nitrite, compared with other globins. This may be due to the positively charged Arg-72(E10) residue, which may help in facilitating electron transfer from the heme to the bound nitrite. The relatively fast conversion of nitrite to NO mediated by the ferrous protein could also be involved in efficient neuronal signaling and affect local activities of nearby neurons. Moreover, it has been shown that NO mediates the inhibitory effects of an FMRF-amide-like neuropeptide in *A. suum*, linking NO with neuropeptide signaling (93). Clearly, the effects of such reaction on the effector domain of the full-length protein remain to be investigated, particularly because *C. elegans* does not express nitric-oxide synthase enzymes and could then in principle be more reliant on nitrite reductase activities for NO signaling.

The covalent structural coupling of a globin domain and a 7TM domain is unique and has not been observed before. However, a functional noncovalent interaction has been shown for the O_2 sensor GLB-5, one of the globins encoded in the genome of *C. elegans*, and NPR-1, an FMRF-amide receptor activated by the binding of the neuropeptides FLP-18 and FLP-21 (8, 9, 94). GLB-5 modulates the same O_2-sensing neurons as NPR-1 (AQR, PQR, and URX), which also expresses GCY-35 and GCY-36, subunits of soluble guanylate cyclases that also act as O_2 sensors (9, 95–100). It was proposed that GLB-5 acts as a signaling molecule, inhibiting neuronal activation when the O_2 concentration drops below 21% (9). Interestingly, soluble guanylate cyclases can also be activated by NO, emphasizing a possible similar mechanism between GLB-33 and GLB-5 (83). Because of its very high affinity for O_2, GLB-33 may function as a highly sensitive O_2 sensor that would modulate neuronal sensitivity to neuropeptides and allow *C. elegans* to move toward nutrient-rich substrates in the hypoxic soil. Furthermore, the ability of GLB-33 GD to generate NO from nitrite (largely available in the soil) under hypoxic conditions would then add another level of complexity to this response.

CONCLUSION

Overall, we can conclude that GLB-33 is composed of a 7TM domain, similar to FMRF-amide receptor domains, covalently coupled to a globin domain that is able to bind external ligands at the heme group. Our modeling data show that the conformational switch of GLB-33 7TM upon ligand binding, likely a neuropeptide ligand, can result in a conformational change of the globin domain and conversely that reactions in the heme pocket may modulate the affinity of neuropeptides for the 7TM domain.

Characterization of GLB-33 GD showed that it possesses characteristics typical of an O_2 sensor when present in the ferrous and ferric forms. Furthermore, the ferric heme occurs as a combination of two hydroxide-ligated ferric forms under physiological conditions. In addition, because nitrite reductase activity is fast compared with that of other globins, GLB-33 GD could also be involved in local NO production from available nitrite when in the ferrous unliganded state, provided that close targets of NO bioactivity and heme-reducing pathways are present. We identify Arg-72(E10) as the most likely trigger for a conformational switch that could be transmitted to the 7TM domain and as such activate yet unknown neuronal signaling pathways.

Acknowledgments—We thank Prof. P. Geldhof and Prof. E. Claerebout (Ghent University, Belgium) for providing A. suum cDNA. We thank the reviewers for the constructive criticism.

REFERENCES

1. Baumgarthl, H., Kritzler, K., Zimelka, W., and Zinkler, D. (1994) Local PO_2 measurements in the environment of submerged soil microarthroods. *Acta Oecologica* **15,** 781–789
2. Hoogewijs, D., Geuens, E., Dewilde, S., Moens, L., Vierstraete, A., Vinogradov, S., and Vanfleteren, J. (2004) Genome-wide analysis of the globin gene family of *C. elegans*. *IUBMB Life* **56,** 697–702
3. Hoogewijs, D., Geuens, E., Dewilde, S., Vierstraete, A., Moens, L., Vinogradov, S., and Vanfleteren, J. R. (2007) Wide diversity in structure and expression profiles among members of the *Caenorhabditis elegans* globin protein family. *BMC Genomics* **8,** 356
4. Hoogewijs, D., De Henau, S., Dewilde, S., Moens, L., Couvreur, M., Borgonie, G., Vinogradov, S. N., Roy, S. W., and Vanfleteren, J. R. (2008) The *Caenorhabditis* globin gene family reveals extensive nematode-specific radiation and diversification. *BMC. Evol. Biol.* **8,** 279
5. Tilleman, L., Germani, F., De Henau, S., Geuens, E., Hoogewijs, D., Braeckman, B. P., Vanfleteren, J. R., Moens, L., and Dewilde, S. (2011) Globins in *Caenorhabditis elegans*. *IUBMB Life* **63,** 166–174
6. Geuens, E., Hoogewijs, D., Nardini, M., Vinck, E., Pesce, A., Kiger, L., Fago, A., Tilleman, L., De Henau, S., Marden, M. C., Weber, R. E., Van Doorslaer, S., Vanfleteren, J., Moens, L., Bolognesi, M., and Dewilde, S. (2010) Globin-like proteins in *Caenorhabditis elegans: in vivo* localization, ligand binding and structural properties. *BMC Biochem.* **11,** 17
7. Yoon, J., Herzik, M. A., Jr., Winter, M. B., Tran, R., Olea, C., Jr., and Marletta, M. A. (2010) Structure and properties of a bis-histidyl ligated globin from *Caenorhabditis elegans*. *Biochemistry* **49,** 5662–5670
8. McGrath, P. T., Rockman, M. V., Zimmer, M., Jang, H., Macosko, E. Z., Kruglyak, L., and Bargmann, C. I. (2009) Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors. *Neuron* **61,** 692–699
9. Persson, A., Gross, E., Laurent, P., Busch, K. E., Bretes, H., and de Bono, M. (2009) Natural variation in a neural globin tunes oxygen sensing in wild *Caenorhabditis elegans*. *Nature* **458,** 1030–1033
10. Tilleman, L., De Henau, S., Pauwels, M., Nagy, N., Pintelon, I., Braeckman, B. P., De Wael, K., Van Doorslaer, S., Adriaensen, D., Timmermans, J. P., Moens, L., and Dewilde, S. (2012) An *N*-myristoylated globin with a redox-sensing function that regulates the defecation cycle in *Caenorhabditis elegans*. *PLoS One* **7,** e48768
11. Casey, P. J., and Gilman, A. G. (1988) G protein involvement in receptor-effector coupling. *J. Biol. Chem.* **263,** 2577–2580
12. Birnbaumer, L., and Brown, A. M. (1990) G proteins and the mechanism of action of hormones, neurotransmitters, and autocrine and paracrine regulatory factors. *Am. Rev. Respir. Dis.* **141,** S106–S114
13. Attwood, T. K., and Findlay, J. B. (1994) Fingerprinting G-protein-coupled receptors. *Protein Eng.* **7,** 195–203
14. Li, C., Nelson, L. S., Kim, K., Nathoo, A., and Hart, A. C. (1999) Neuropeptide gene families in the nematode *Caenorhabditis elegans*. *Ann. N.Y. Acad. Sci.* **897,** 239–252
15. Li, C. (2005) The ever-expanding neuropeptide gene families in the nematode *Caenorhabditis elegans*. *Parasitology* **131,** S109–S127
16. Li, C., and Kim, K. (2008) Neuropeptides. *WormBook*, 1–36

17. Nathoo, A. N., Moeller, R. A., Westlund, B. A., and Hart, A. C. (2001) Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 14000–14005
18. Husson, S. J., Mertens, I., Janssen, T., Lindemans, M., and Schoofs, L. (2007) Neuropeptidergic signaling in the nematode *Caenorhabditis elegans*. *Prog. Neurobiol.* **82,** 33–55
19. Husson, S. J., Clynen, E., Baggerman, G., De Loof, A., and Schoofs, L. (2005) Discovering neuropeptides in *Caenorhabditis elegans* by two-dimensional liquid chromatography and mass spectrometry. *Biochem. Biophys. Res. Commun.* **335,** 76–86
20. Pierce, S. B., Costa, M., Wisotzkey, R., Devadhar, S., Homburger, S. A., Buchman, A. R., Ferguson, K. C., Heller, J., Platt, D. M., Pasquinelli, A. A., Liu, L. X., Doberstein, S. K., and Ruvkun, G. (2001) Regulation of DAF-2 receptor signaling by human insulin and ins-1, a member of the unusually large and diverse *C. elegans* insulin gene family. *Genes Dev.* **15,** 672–686
21. Li, W., Kennedy, S. G., and Ruvkun, G. (2003) daf-28 encodes a *C. elegans* insulin superfamily member that is regulated by environmental cues and acts in the DAF-2 signaling pathway. *Genes Dev.* **17,** 844–858
22. Li, C., Kim, K., and Nelson, L. S. (1999) FMRF-amide-related neuropeptide gene family in *Caenorhabditis elegans*. *Brain Res.* **848,** 26–34
23. de Bono, M., and Bargmann, C. I. (1998) Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* **94,** 679–689
24. Proulx, K., Richard, D., and Walker, C. D. (2002) Leptin regulates appetite-related neuropeptides in the hypothalamus of developing rats without affecting food intake. *Endocrinology* **143,** 4683–4692
25. Janssen, T., Meelkop, E., Lindemans, M., Verstraelen, K., Husson, S. J., Temmerman, L., Nachman, R. J., and Schoofs, L. (2008) Discovery of a cholecystokinin-gastrin-like signaling system in nematodes. *Endocrinology* **149,** 2826–2839
26. Cowden, C., Stretton, A. O., and Davis, R. E. (1989) AF1, a sequenced bioactive neuropeptide isolated from the nematode *Ascaris suum*. *Neuron* **2,** 1465–1473
27. Davis, R. E., and Stretton, A. O. (2001) Structure-activity relationships of 18 endogenous neuropeptides on the motor nervous system of the nematode *Ascaris suum*. *Peptides* **22,** 7–23
28. Hu, Z., Pym, E. C., Babu, K., Vashlishan Murray, A. B., and Kaplan, J. M. (2011) A neuropeptide-mediated stretch response links muscle contraction to changes in neurotransmitter release. *Neuron* **71,** 92–102
29. Meelkop, E., Temmerman, L., Janssen, T., Suetens, N., Beets, I., Van Rompay, L., Shanmugam, N., Husson, S. J., and Schoofs, L. (2012) PDF receptor signaling in *Caenorhabditis elegans* modulates locomotion and egg-laying. *Mol. Cell. Endocrinol.* **361,** 232–240
30. Janssen, T., Husson, S. J., Lindemans, M., Mertens, I., Rademakers, S., Ver Donck, K., Geysen, J., Jansen, G., and Schoofs, L. (2008) Functional characterization of three G protein-coupled receptors for pigment dispersing factors in *Caenorhabditis elegans*. *J. Biol. Chem.* **283,** 15241–15249
31. Lindemans, M., Liu, F., Janssen, T., Husson, S. J., Mertens, I., Gäde, G., and Schoofs, L. (2009) Adipokinetic hormone signaling through the gonadotropin-releasing hormone receptor modulates egg-laying in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 1642–1647
32. Fierro-González, J. C., Cornils, A., Alcedo, J., Miranda-Vizuete, A., and Swoboda, P. (2011) The thioredoxin TRX-1 modulates the function of the insulin-like neuropeptide DAF-28 during dauer formation in *Caenorhabditis elegans*. *PLoS One* **6,** e16561
33. Bashford, D., Chothia, C., and Lesk, A. M. (1987) Determinants of a protein fold. Unique features of the globin amino acid sequences. *J. Mol. Biol.* **196,** 199–216
34. Moens, L., Vanfleteren, J., Van de Peer, Y., Peeters, K., Kapp, O., Czeluzniak, J., Goodman, M., Blaxter, M., and Vinogradov, S. (1996) Globins in nonvertebrate species: dispersal by horizontal gene transfer and evolution of the structure-function relationships. *Mol. Biol. Evol.* **13,** 324–333
35. Hobert, O. (2002) PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *BioTechniques* **32,** 728–730
36. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402
37. Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A., and Yu, Y. K. (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **272,** 5101–5109
38. Cantacessi, C., Zou, F. C., Hall, R. S., Zhong, W., Jex, A. R., Campbell, B. E., Ranganathan, S., Sternberg, P. W., Zhu, X. Q., and Gasser, R. B. (2009) Bioinformatic analysis of abundant, gender-enriched transcripts of adult *Ascaris suum* (Nematoda) using a semi-automated workflow platform. *Mol. Cell. Probes* **23,** 205–217
39. Martin, J., Abubucker, S., Wylie, T., Yin, Y., Wang, Z., and Mitreva, M. (2009) Nematode.net update 2008: improvements enabling more efficient data mining and comparative nematode genomics. *Nucleic Acids Res.* **37,** D571–D578
40. Yook, K., Harris, T. W., Bieri, T., Cabunoc, A., Chan, J., Chen, W. J., Davis, P., de la Cruz, N., Duong, A., Fang, R., Ganesan, U., Grove, C., Howe, K., Kadam, S., Kishore, R., *et al.* (2012) WormBase 2012: more genomes, more data, new website. *Nucleic Acids Res.* **40,** D735–D741
41. Nicholas, K. B., Nicholas, H. B. Jr., and Deerfield, D. W., II. (1997) GeneDoc: analysis and visualization of genetic variation. *EMBNEW NEWS* **4,** 14
42. Vinogradov, S. N., Hoogewijs, D., Bailly, X., Arredondo-Peter, R., Guertin, M., Gough, J., Dewilde, S., Moens, L., and Vanfleteren, J. R. (2005) Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 11385–11389
43. Vinogradov, S. N., Hoogewijs, D., Bailly, X., Arredondo-Peter, R., Gough, J., Dewilde, S., Moens, L., and Vanfleteren, J. R. (2006) A phylogenomic profile of globins. *BMC Evol. Biol.* **6,** 31
44. Dewilde, S., Kiger, L., Burmester, T., Hankeln, T., Baudin-Creuza, V., Aerts, T., Marden, M. C., Caubergs, R., and Moens, L. (2001) Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family. *J. Biol. Chem.* **276,** 38949–38955
45. Antonini, E., and Brunori, M. (1971) in *Hemoglobin and Myoglobin in Their Recation with Ligands* (Neuberger, A. and Tatum, E. L, eds) North Holland Publishing, Amsterdam
46. Sick, H., and Gersonde, K. (1969) Method for continuous registration of O2-binding curves of hemoproteins by means of a diffusion chamber. *Anal. Biochem.* **32,** 362–376
47. Weber, R. E. (1992) Use of ionic and zwitterionic (Tris Bistris and Hepes) buffers in studies on hemoglobin-function. *J. Appl. Physiol.* **72,** 1611–1615
48. Helbo, S., and Fago, A. (2011) Allosteric modulation by *S*-nitrosation in the low O2 affinity myoglobin from rainbow trout. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **300,** R101–R108
49. Schmidt, H. H., and Kelm, M. (1996) *Methods in Nitric Oxide Research* (Feelisch, M., and Stamler, J. S. eds) pp. 491–497, Wiley & Sons Ltd., Chichester, UK
50. Salhany, J. M. (2008) Kinetics of reaction of nitrite with deoxy hemoglobin after rapid deoxygenation or predeoxygenation by dithionite measured in solution and bound to the cytoplasmic domain of band 3 (SLC4A1). *Biochemistry* **47,** 6059–6072
51. Pedersen, C. L., Faggiano, S., Helbo, S., Gesser, H., and Fago, A. (2010) Roles of nitric oxide, nitrite and myoglobin on myocardial efficiency in trout (*Oncorhynchus mykiss*) and goldfish (*Carassius auratus*): implications for hypoxia tolerance. *J. Exp. Biol.* **213,** 2755–2762
52. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9,** 40
53. Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5,** 725–738
54. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4,** 363–371
55. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) Stereochemical quality of protein structure coordinates. *Proteins* **12,** 345–364
56. Bowie, J. U., Lüthy, R., and Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure.

Science **253,** 164–170
57. Wiederstein, M., and Sippl, M. J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35,** W407–W410
58. Delano, W. L. (2002) *The PyMOL Molecular Graphics System,* DeLano Scientific LLC, San Carlos, CA
59. Lyskov, S., and Gray, J. J. (2008) The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **36,** W233–W238
60. Mitreva, M., Blaxter, M. L., Bird, D. M., and McCarter, J. P. (2005) Comparative genomics of nematodes. *Trends Genet.* **21,** 573–581
61. Hanania, G. I., Yeghiayan, A., and Cameron, B. F. (1966) Absorption spectra of sperm-whale ferrimyoglobin. *Biochem. J.* **98,** 189–192
62. Yoshida, T., and Migita, C. T. (2000) Mechanism of heme degradation by heme oxygenase. *J. Inorg. Biochem.* **82,** 33–41
63. Ratliff, M., Zhu, W., Deshmukh, R., Wilks, A., and Stojiljkovic, I. (2001) Homologues of neisserial heme oxygenase in Gram-negative bacteria: degradation of heme by the product of the *pigA* gene of *Pseudomonas aeruginosa*. *J. Bacteriol.* **183,** 6394–6403
64. Couture, M., Adak, S., Stuehr, D. J., and Rousseau, D. L. (2001) Regulation of the properties of the heme-NO complexes in nitric-oxide synthase by hydrogen bonding to the proximal cysteine. *J. Biol. Chem.* **276,** 38280–38288
65. Cho, K. C., Remba, R. D., and Fitchen, D. B. (1981) Resonance Raman studies of methemoglobin derivatives at room temperature and 77 K. *Biochim. Biophys. Acta* **668,** 186–192
66. Rousseau, D. L., Ching, Y. C., Brunori, M., and Giacometti, G. M. (1989) Axial coordination of ferric *Aplysia* myoglobin. *J. Biol. Chem.* **264,** 7878–7881
67. Takahashi, S., Ishikawa, Takeuchi, N., Ikeda-Saito, M., Yoshida, Y., and Rousseau, D. L. (1995) Oxygen-bound heme-heme oxygenase complex: evidence for a highly bent structure of the coordinated oxygen. *J. Am. Chem. Soc.* **117,** 6002–6006
68. Hu, S, Z., Smith, K., and Spiro, T. G. (1996) Assignment of protoheme resonance raman spectrum by heme labeling in myoglobin. *J. Am. Chem. Soc.* **118,** 12638–12646
69. Das, T. K., Lee, H. C., Duff, S. M., Hill, R. D., Peisach, J., Rousseau, D. L., Wittenberg, B. A., and Wittenberg, J. B. (1999) The heme environment in barley hemoglobin. *J. Biol. Chem.* **274,** 4207–4212
70. Aono, S., Kato, T., Matsuki, M., Nakajima, H., Ohta, T., Uchida, T., and Kitagawa, T. (2002) Resonance Raman and ligand binding studies of the oxygen-sensing signal transducer protein HemAT from *Bacillus subtilis*. *J. Biol. Chem.* **277,** 13528–13538
71. Sitter, A. J., Shifflett, J. R., and Terner, J. (1988) Resonance Raman spectroscopic evidence for heme iron-hydroxide ligation in peroxidase alkaline forms. *J. Biol. Chem.* **263,** 13032–13038
72. Spiro, T. G., and Wasbotten, I. H. (2005) CO as a vibrational probe of heme protein active sites. *J. Inorg. Biochem.* **99,** 34–44
73. Vergara, A., Franzese, M., Merlino, A., Vitagliano, L., Verde, C., di Prisco, G., Lee, H. C., Peisach, J., and Mazzarella, L. (2007) Structural characterization of ferric hemoglobins from three antarctic fish species of the suborder notothenioidei. *Biophys. J* **93,** 2822–2829
74. Blumberg, W. E., and Peisach, J. (1972) Low-spin ferric forms of hemoglobin and other heme proteins. *Wenner-Gren Center International Symposium Series* **18,** 219–225
75. Wittenberg, B. A., Kampa, L., Wittenberg, J. B., Blumberg, W. E., and Peisach, J. (1968) The electronic structure of protoheme proteins. II. An electron paramagnetic resonance and optical study of cytochrome c peroxidase and its derivatives. *J. Biol. Chem.* **243,** 1863–1870
76. Hori, H., and Yonetani, T. (1985) Powder and single-crystal electron paramagnetic resonance studies of yeast cytochrome c peroxidase and its peroxide and its peroxide compound, Compound ES. *J. Biol. Chem.* **260,** 349–355
77. Blumberg, W. E., Peisach, J., Wittenberg, B. A., and Wittenberg, J. B. (1968) The electronic structure of protoheme proteins. I. An electron paramagnetic resonance and optical study of horseradish peroxidase and its derivatives. *J. Biol. Chem.* **243,** 1854–1862
78. Foote, N., Gadsby, P. M., Berry, M. J., Greenwood, C., and Thomson, A. J. (1987) The formation of ferric haem during low temperature photolysis of horseradish peroxidase Compound I. *Biochem. J.* **246,** 659–668
79. Nicoletti, F. P., Droghetti, E., Howes, B. D., Bustamante, J. P., Bonamore, A., Sciamanna, N., Estrin, D. A., Feis, A., Boffi, A., and Smulevich, G. (2013) H-bonding networks of the distal residues and water molecules in the active site of *Thermobifida fusca* hemoglobin. *Biochim. Biophys. Acta* **1834,** 1901–1909
80. Boutilier, R. G., Heming, T. A., and Iwama, G. K. (1984) Appendix–physicochemical parameters for use in fish respiratory physiology. *Fish Physiol.* **10,** 403–430
81. Suzuki, T., and Imai, K. (1998) Evolution of myoglobin. *Cell. Mol. Life Sci.* **54,** 979–1004
82. Brantley, R. E., Jr., Smerdon, S. J., Wilkinson, A. J., Singleton, E. W., and Olson, J. S. (1993) The mechanism of autooxidation of myoglobin. *J. Biol. Chem.* **268,** 6995–7010
83. Shikama, K. (1998) The molecular mechanism of autoxidation for myoglobin and hemoglobin: a venerable puzzle. *Chem. Rev.* **98,** 1357–1374
84. Shiva, S., Huang, Z., Grubina, R., Sun, J., Ringwood, L. A., MacArthur, P. H., Xu, X., Murphy, E., Darley-Usmar, V. M., and Gladwin, M. T. (2007) Deoxymyoglobin is a nitrite reductase that generates nitric oxide and regulates mitochondrial respiration. *Circ. Res.* **100,** 654–661
85. Tiso, M., Tejero, J., Basu, S., Azarov, I., Wang, X., Simplaceanu, V., Frizzell, S., Jayaraman, T., Geary, L., Shapiro, C., Ho, C., Shiva, S., Kim-Shapiro, D. B., and Gladwin, M. T. (2011) Human neuroglobin functions as a redox-regulated nitrite reductase. *J. Biol. Chem.* **286,** 18277–18289
86. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K., and Stevens, R. C. (2007) High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. *Science* **318,** 1258–1265
87. Travaglini Allocatelli, C., Cutruzzolà, F., Brancaccio, A., Vallone, B., and Brunori, M. (1994) Engineering *Ascaris* hemoglobin oxygen affinity in sperm whale myoglobin: role of tyrosine B10. *FEBS Lett.* **352,** 63–66
88. Pesce, A., Dewilde, S., Kiger, L., Milani, M., Ascenzi, P., Marden, M. C., Van Hauwaert, M. L., Vanfleteren, J., Moens, L., and Bolognesi, M. (2001) Very high resolution structure of a trematode hemoglobin displaying a TyrB10-TyrE7 heme distal residue pair and high oxygen affinity. *J. Mol. Biol.* **309,** 1153–1164
89. Allocatelli, C. T., Cutruzzolà, F., Brancaccio, A., Brunori, M., Qin, J., and La Mar, G. N. (1993) Structural and functional characterization of sperm whale myoglobin mutants: role of arginine (E10) in ligand stabilization. *Biochemistry* **32,** 6041–6049
90. Howes, B. D., Giordano, D., Boechi, L., Russo, R., Mucciacciaro, S., Ciaccio, C., Sinibaldi, F., Fittipaldi, M., Martí, M. A., Estrin, D. A., di Prisco, G., Coletta, M., Verde, C., and Smulevich, G. (2011) The peculiar heme pocket of the 2/2 hemoglobin of cold-adapted *Pseudoalteromonas haloplanktis* TAC125. *J. Biol. Inorg. Chem.* **16,** 299–311
91. Cutruzzolà, F., Travaglini Allocatelli, C., Brancaccio, A., and Brunori, M. (1996) *Aplysia limacina* myoglobin cDNA cloning: an alternative mechanism of oxygen stabilization as studied by active-site mutagenesis. *Biochem. J.* **314,** 83–90
92. Dunham, C. M., Dioum, E. M., Tuckerman, J. R., Gonzalez, G., Scott, W. G., and Gilles-Gonzalez, M. A. (2003) A distal arginine in oxygen-sensing heme-PAS domains is essential to ligand binding, signal transduction, and structure. *Biochemistry* **42,** 7701–7708
93. Bowman, J. W., Winterrowd, C. A., Friedman, A. R., Thompson, D. P., Klein, R. D., Davis, J. P., Maule, A. G., Blair, K. L., and Geary, T. G. (1995) Nitric oxide mediates the inhibitory effects of SDPNFLRFamide, a nematode FMRF-amide-related neuropeptide, in Ascaris suum. *J. Neurophysiol.* **74,** 1880–1888
94. Rogers, C., Reale, V., Kim, K., Chatwin, H., Li, C., Evans, P., and de Bono, M. (2003) Inhibition of *Caenorhabditis elegans* social feeding by FMRF-amide-related peptide activation of NPR-1. *Nat. Neurosci.* **6,** 1178–1185
95. Cheung, B. H., Arellano-Carbajal, F., Rybicki, I., and de Bono, M. (2004) Soluble guanylate cyclases act in neurons exposed to the body fluid to promote *C. elegans* aggregation behavior. *Curr. Biol.* **14,** 1105–1111
96. Gray, J. M., Karow, D. S., Lu, H., Chang, A. J., Chang, J. S., Ellis, R. E., Marletta, M. A., and Bargmann, C. I. (2004) Oxygen sensation and social

feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature* **430,** 317–322
97. Cheung, B. H., Cohen, M., Rogers, C., Albayram, O., and de Bono, M. (2005) Experience-dependent modulation of *C. elegans* behavior by ambient oxygen. *Curr. Biol.* **15,** 905–917
98. Yu, S., Avery, L., Baude, E., and Garbers, D. L. (1997) Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 3384–3387
99. Rogers, C., Persson, A., Cheung, B., and de Bono, M. (2006) Behavioral motifs and neural pathways coordinating O2 responses and aggregation in *C. elegans*. *Curr. Biol.* **16,** 649–659
100. Zimmer, M., Gray, J. M., Pokala, N., Chang, A. J., Karow, D. S., Marletta, M. A., Hudson, M. L., Morton, D. B., Chronis, N., and Bargmann, C. I. (2009) Neurons detect increases and decreases in oxygen levels using distinct guanylate cyclases. *Neuron* **61,** 865–879
101. Olson, J. S., and Philips, G. N. (1997) Myoglobin discriminates between oxygen, NO, and CO by electrostatic interactions with the bound ligand. *J. Biol. Inorg. Chem.* **2,** 544–552
102. De Baere, I., Perutz, M.F., Kiger, L., Marden, M.C., and Poyart, C. (1994) Formation of two hydrogen bonds from the globin to the heme-linked oxygen molecule in *Ascaris* hemoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 1594–1597

A Globin Domain in a Neuronal Transmembrane Receptor of *Caenorhabditis elegans* and *Ascaris suum*: MOLECULAR MODELING AND FUNCTIONAL PROPERTIES

Lesley Tilleman, Francesca Germani, Sasha De Henau, Signe Helbo, Filip Desmet, Herald Berghmans, Sabine Van Doorslaer, David Hoogewijs, Liliane Schoofs, Bart P. Braeckman, Luc Moens, Angela Fago and Sylvia Dewilde

J. Biol. Chem. 2015, 290:10336-10352.
doi: 10.1074/jbc.M114.576520 originally published online February 9, 2015
